DYL-Leaf: A Lightweight Distilled YOLO-based Model for Plant Leaf Disease Classification

Touhid Alam
Department of Computer Science
American International University-Bangladesh
1229, Dhaka, Bangladesh
22-46330-1@student.aiub.edu

Abir Bokhtiar
Department of Computer Science
American International University-Bangladesh
1229, Dhaka, Bangladesh
22-47038-1@student.aiub.edu

Md. Saef Ullah Miah
Department of Computer Science
American International University-Bangladesh
1229, Dhaka, Bangladesh
saef@aiub.edu

Junaida Sulaiman*
Faculty of Computing
Universiti Malaysia Pahang Al-Sultan Abdullah
26600 Pekan, Pahang, Malaysia
junaida@umpsa.edu.my

2025 IEEE 9th International Conference on Software Engineering & Computer Systems (ICSECS) | 979-8-3315-4441-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICSECS65227.2025.11279233

***Abstract*—Agriculture is vital for global food security, but plant leaf diseases pose escalating threats, causing significant crop losses and economic damage. Traditional diagnostic methods are often time-consuming and resource-intensive, prompting the need for efficient, scalable solutions. This research addresses these challenges by proposing DYL-Leaf, a lightweight, distilled model designed for detecting 13 classes of potato, rice, and tomato leaf diseases from the PlantVillage dataset. Leveraging Knowledge Distillation (KD), a lightweight student model with only 545,005 parameters, is trained to emulate a larger, custom YOLO-based teacher model (2.6M parameters). The methodology optimizes both hard and soft losses using temperature-scaled Kullback-Leibler divergence, ensuring the student model retains the teacher's knowledge while being computationally efficient. Results demonstrate that the student model not only matches but surpasses the teacher's performance, achieving a validation accuracy of 93.8% (vs. 92.9%), along with improved precision (94.00%), recall (93.23%), and F1-score (93.38%). Saliency maps were employed to interpret the model's decision-making process, confirming its ability to focus on disease-specific features. These findings highlight the effectiveness of KD in creating lightweight, high-performing models for plant disease classification. By reducing computational requirements while maintaining accuracy, DYL-Leaf is well-suited for deployment in resource-constrained agricultural environments.**

***Keywords*—Computer Vision, Knowledge Distillation, Plant Disease Classification, YOLO**

I. Introduction

The global population is projected to surpass 9 billion by 2050, intensifying pressure on agriculture to meet rising food demands [1]. Studies estimate a 35%–56% increase in food demand by 2050, while food insecurity remains a critical concern [2]. Among various challenges, leaf diseases significantly threaten crop yields, causing up to 40% of global agricultural losses annually, which translates to an estimated economic impact of over $220 billion [3]. These diseases, driven by fungal, bacterial, and viral infections [4], have led to severe outbreaks, particularly in Africa and Asia [5]. For instance, blight and blast diseases severely impact tomato, potato, and rice crops, leading to economic losses and food security risks [6], [7]. Early and accurate classification of leaf diseases is crucial for mitigating crop losses. Traditional diagnostic methods, such as visual inspection and microscopy, are often inefficient and require specialized expertise [8]. Recent advancements in Computer Vision (CV) have revolutionized plant disease classification, with You Only Look Once (YOLO) based models demonstrating strong performance due to their balance of speed and accuracy [9]. Modern YOLO architecture continues to improve feature extraction and model efficiency.

This study employs Knowledge Distillation (KD) to develop an efficient model for leaf disease detection. KD enables a lightweight "student" model to learn from a larger "teacher" model [10], enhancing efficiency while maintaining high accuracy. Our proposed model, DYL-Leaf, contains only 545,005 parameters—significantly fewer than its 2.6 million parameter teachers while effectively classifying 13 leaf disease categories in potato, rice, and tomato using 4,144 images from the PlantVillage dataset.

II. Literature Review

Recent advancements in CV have significantly improved plant disease classification, addressing the growing need for efficient and scalable solutions in agriculture. Yakkundimath et al. [11] detected rice leaf disease using a pre- trained Convolutional Neural Network (CNN), achieving an accuracy of 92.24% with VGG16 [11]. Their dataset contained 1,200 images, augmented to 12,000 samples. H. Yin et al. used ResNet50 to detect hot pepper leaf diseases with 88.38% accuracy, enhancing their model with a KNN-based deep feature extraction system [12]. Y. Kurmi et al. developed a custom CNN model that achieved 95.8% accuracy for bell pepper, 94.1% for potato, and 92.6% for tomato leaf disease classification on the PlantVillage dataset [13]. S. S. Gaikwad et al. compared AlexNet and SqueezeNet for fruit leaf disease classification, with AlexNet reaching 86.8% accuracy [14]. M. Prabu et al. combined MobileNetV2 with a Support Vector Machine (SVM) to detect Mango leaf diseases, reporting 94.5% accuracy [15].

While these prior works demonstrate high accuracy, they often rely on large models such as VGG16 and ResNet50, which contain tens of millions of parameters. This large model size renders them computationally expensive and impractical for on-device deployment in resource-constrained agricultural settings.

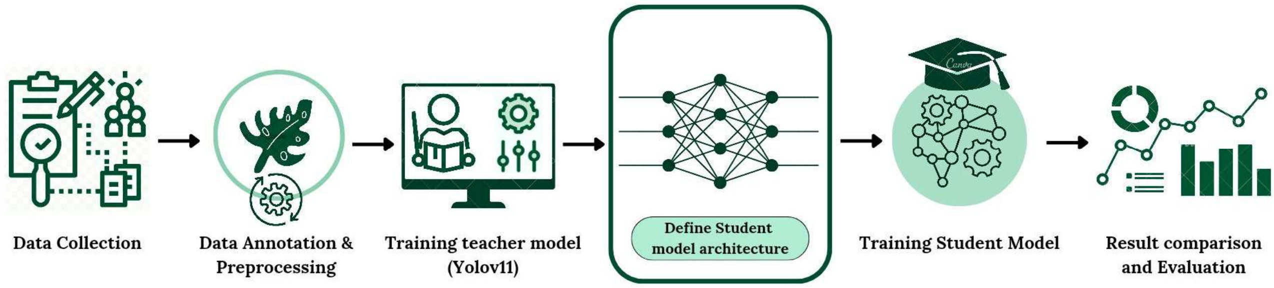

Fig. 1. Workflow of the proposed methodology of the study.

This research addresses this critical gap by employing knowledge distillation to develop DYL-Leaf, a highly compact model with only 545,005 parameters that not only maintains but exceeds the performance of larger architectures, making it suitable for real-world application.

III. Methodology

This section outlines the methodology for developing DYL-Leaf, a lightweight distilled YOLO-based model for plant leaf disease classification. The workflow, illustrated in Fig. 1, includes data collection, preprocessing, model architecture, and knowledge distillation

A. Data Collection and Preporcessing

For this study, a targeted subset of the PlantVillage dataset [16] was used, containing 4144 images across 13 classes of potato, rice, and tomato leaf diseases (Table I). These crops were selected due to their economic significance and the prevalence of the chosen diseases. This focused approach allowed for a controlled experiment on a balanced and relevant dataset. Data preprocessing techniques, including resizing (256x256), random flipping, rotation (10°), brightness/contrast adjustment, grayscale conversion, and normalization, were applied to enhance model generalization (Fig. 2 and Table II).

B. Teacher Model Architecture

The teacher model is a custom YOLO-based architecture with 2.6M parameters, comprising three main modules:

*1) **Backbone:***

Extracts features using convolutional layers and C3k2 blocks, down sampling the input to an 8x8x128 feature map.

*2) **Neck:***

Enhances features using C2PSA and SPPF blocks, focusing on small object details and aggregating multi-scale features.

*3) **Head:***

Performs classification and localization at multiple scales (80x80, 40x40, 20x20).

Key advancements incorporated in this architecture include the C3k2 block for efficient feature extraction with fewer parameters, the C2PSA (Channel and Spatial Attention) block for improved spatial attention, and the SPPF (Spatial Pyramid Pooling Fast) block for robust multi-scale feature fusion (Fig. 3).

TABLE I. Class Name And Instance Of The Dataset

Class Name	Instances
Potato Early Blight	223
Potato Late Blight	353
Potato healthy	152
Rice Bacterialblight	339
Rice Blast	416
Rice Blight	213
Rice Brownspot	317
Rice Tungro	260
Tomato Late Blight	407
Tomato Spider Mites	324
Tomato Yellow Leaf Curl Virus	568
Tomato Mosaic Virus	373
Tomato Healthy	199
Total	**4144**

TABLE II. Data Preprocessing Techniques

Preprocessing Technique	Value
Resize	256x256
Horizontal Flip	50% probability
Random Rotation	10 degrees
Brightness Adjustment	0.2
Contrast Adjustment	0.2
Random Grayscale	10% probability
Normalization	[0, 1]

C. Proposed Student Model (DYL-Leaf)

The student model, DYL-Leaf, shown in Fig. 4, is a significantly streamlined architecture with only 545,005 parameters. It consists of:

a) Backbone:

Uses a DarkNet-inspired structure with C3k2 and SPPF for efficient feature extraction. The input of 256x256x3 is down sampled to 8x8x128.

b) Neck:

Employs a lightweight DarkFPN for multi-scale feature aggregation, preparing the features for the final classification task.

c) Head:

A simple classifier head that processes the aggregated features from the neck to predict one of the 13 disease classes.

The compact design of DYL-Leaf makes it ideal for deployment on resource-limited devices.

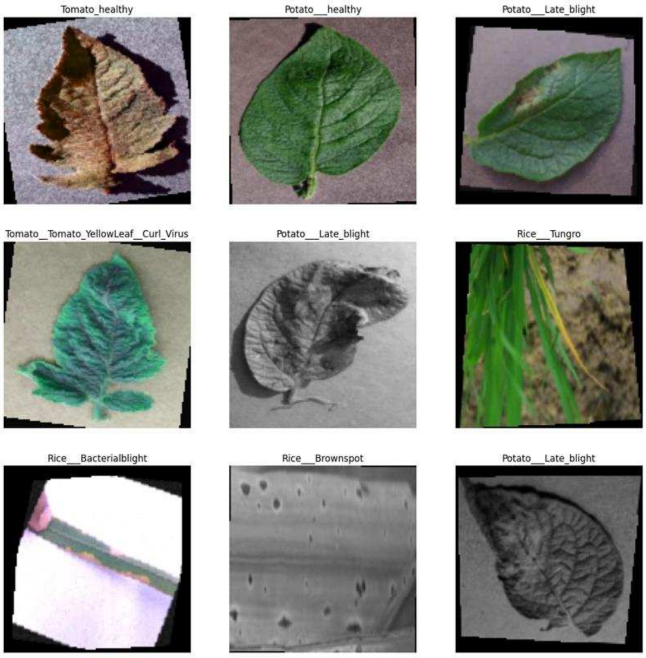

Fig 2. Example of the classes of the dataset after data preprocessing.

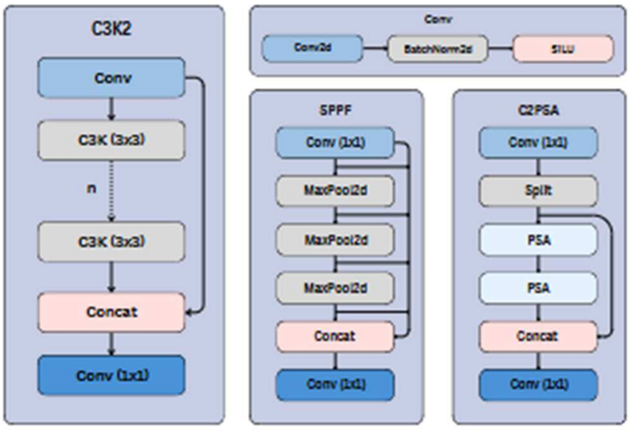

Fig. 3. Architectural details of the custom blocks used in the teacher model.

D. Knowledge Distillation (KD):

KD transfers knowledge from the larger teacher to the smaller student model by minimizing a combined loss function in (1):

$$L_{distll} = \alpha * L_{hard} + (1 - \alpha) * L_{soft}. \quad (1)$$

where L_{hard} is the standard cross-entropy loss against the ground-truth labels. L_{soft} is the Kullback-Leibler (KL) divergence between the softened outputs of the teacher and student models, calculated with a temperature scaling factor T. Following common practices in KD literature [10], we set the temperature to T = 2 to soften the probability distributions, which helps the student learn inter-class relationships from the teacher's soft label.

E. Training Specifications:

The model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32. Early stopping with a patience of 30 epochs was employed to prevent overfitting. The training was conducted for a maximum of 100 epochs.

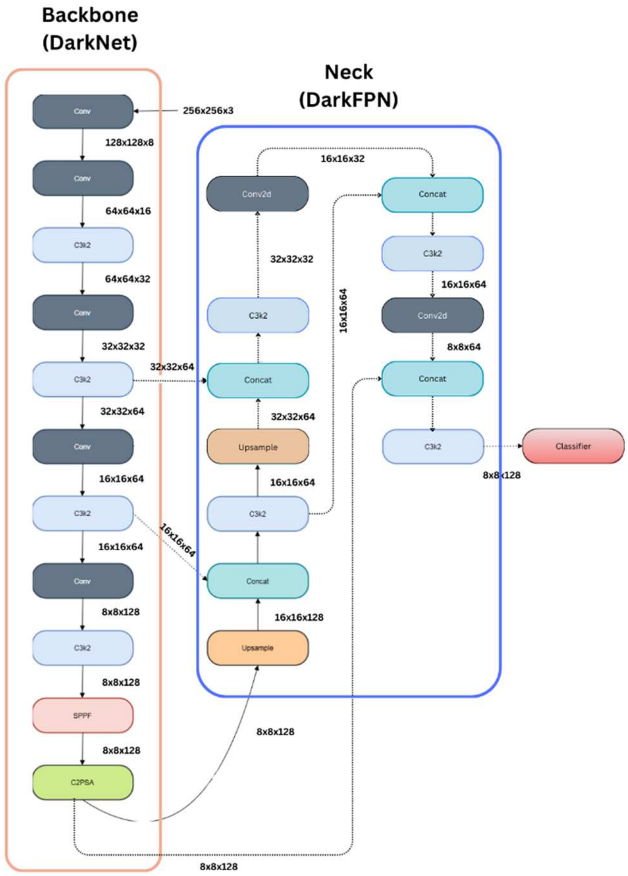

Fig. 4. Architecture of proposed student model, DYL-Leaf.

F. Evalution Matrics:

The model's performance was assessed using standard evaluation metrics, including training and validation accuracy, precision, recall, F1-score, and a confusion matrix to analyze class-wise performance.

G. Model Interpretation:

Saliency maps were generated to interpret the model's decision-making process by visualizing the input image regions that most significantly influenced the final prediction. This helps to ensure the model focuses on relevant pathological features.

IV. Result and Discussion

This section presents a comprehensive analysis of the experimental results, comparing the performance of the proposed student model, DYL-Leaf, against its larger teacher model. The findings demonstrate not only the successful application of knowledge distillation for model compression but also a significant improvement in predictive accuracy and practical deployability. The overall performance comparison, summarized in Table III, reveals that the distilled student model, DYL-Leaf, achieves a validation accuracy of 93.8%, surpassing the 92.9% accuracy of its 2.6M parameter teacher. This superior performance is consistent across all key metrics, with DYL-leaf showing higher precision (94.00% vs. 91.42%), recall (93.23% vs. 90.00%), and F1-score (93.38% vs. 90.42%). This is a critical finding: despite an 80% reduction in trainable parameters (to just 545,005), the student models are not only more efficient but also more effective. The training and validation loss curves provide further insight into model behavior. As seen in Fig. 5, the teacher model shows good generalization. However, the student model's curves in Fig. 6 exhibit a smoother and more stable convergence, suggesting a more robust learning process free from minor fluctuations.

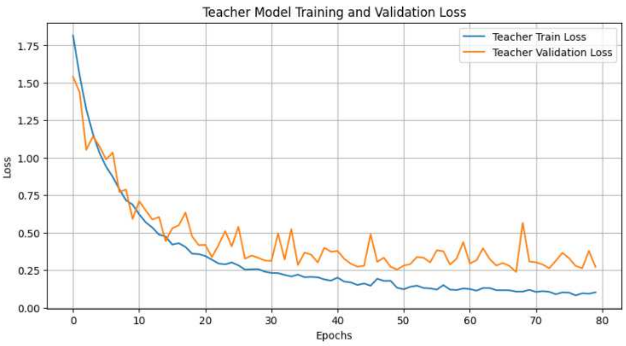

Fig. 5. Training and validation loss of the teacher model.

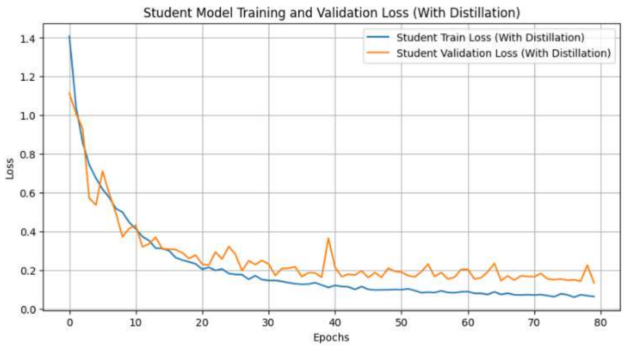

Fig. 6. Training and validation loss of the student model (DYL-Leaf).

The practical significance of the student model outperforming the teacher is substantial. This counter-intuitive result suggests that knowledge distillation can function as a potent form of regularization. The larger teacher model, with its high parametric complexity, may be prone to overfitting by learning spurious correlations or noise from the training data. In contrast, the architecturally constrained student model is forced to learn a more compact and generalized representation. By being guided by the teacher's softened probability distributions, the student distills the most essential knowledge while being inherently robust to noise. From a deployment perspective, this is a best-of-both-worlds scenario: DYL-Leaf offers a higher degree of diagnostic reliability while its minimal computational footprint makes it ideally suited for real-world, on-field applications on resource-constrained devices like smartphones or drones, where both accuracy and efficiency are paramount. A deeper, class-wise performance analysis, detailed in Table IV for the teacher and Table V for the student, reveals the nuances of this improvement. The teacher model struggled with certain classes, such as "Potato healthy" (recall of 0.60) and "Rice Blight" (F1-score of 0.71). DYL-Leaf rectifies these weaknesses, boosting the recall for "Potato healthy" to a robust 0.87 and achieving perfect scores for "Rice Blast" and "Tomato healthy".

Despite these improvements, "Rice Blight" remained the most challenging class. Further exploration suggests the reason may be rooted in the data itself; its visual symptoms can be subtle and highly variable, making it inherently difficult to distinguish. The confusion matrix in Fig. 7 corroborates this, showing a strong diagonal for most classes but with minor confusions linked to these more visually ambiguous diseases. This insight suggests that future work could benefit from incorporating more diverse and expertly annotated examples of this specific class to enhance model robustness. To compare Fig. 8 also shows the confusion matrix of the teacher model.

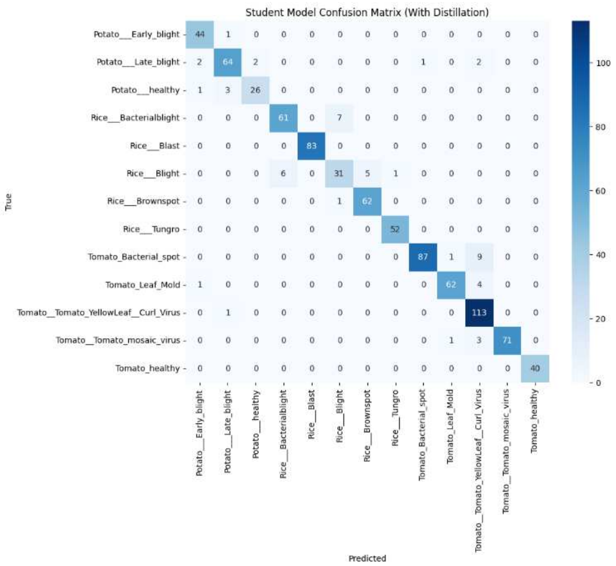

Fig. 7. Confusion matrix of the student model (DYL-Leaf).

TABLE III. TEACHER AND STUDENT MODEL PERFORMANCE COMPARISON

Model	Parameters	Training Accuracy	Validation Accuracy	Precision	Recall	F1 Score
Teacher (YOLOv11)	2.6M	96.4%	92.9%	91.42%	90.00%	90.42%
Student (DYL-Leaf	545,005	96.9%	93.8%	94.00%	93.23%	93.38%

To ensure the model's decisions are trustworthy, we employed saliency maps (Fig. 9) to interpret its predictions. These visualizations confirm that DYL-Leaf's decision-making process is sound, as the highlighted regions consistently correspond to the actual diseased portions of the leaves. This is critical for building a reliable diagnostic tool. Notably, the distilled student model not only matched but surpassed the performance of its larger teacher, achieving a validation accuracy of 93.8% compared to the teacher's 92.9%. This seemingly counter-intuitive outcome suggests that knowledge distillation can function as a potent form of regularization. The teacher model, with its greater parametric complexity (2.6M parameters), possesses a high capacity that may lead it to overfit the training data, learning spurious correlations or noise in addition to the core features of each disease class. The student model, by design, is architecturally constrained with only 545,005 parameters, limiting its ability to memorize such idiosyncrasies.

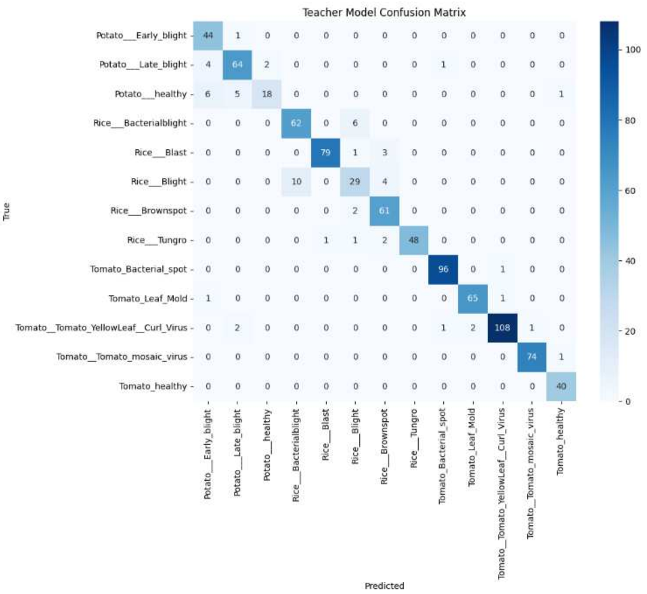

Fig. 8. Confusion matrix of the teacher model.

TABLE IV. PRECISION, RECALL, F1-SCORE AND SUPPORT FOR EACH CLASS OF THE TEACHER MODEL

Class	Precision	Recall	F1 score	Support
Potato Early blight	0.80	0.98	0.88	45
Potato Late blight	0.89	0.90	0.90	71
Potato healthy	0.90	0.60	0.72	30
Rice Bacterialblight	0.86	0.91	0.89	68
Rice Blast	0.99	0.95	0.97	83
Rice Blight	0.74	0.67	0.71	43
Rice Brownspot	0.87	0.97	0.92	63
Rice Tungro	1.00	0.92	0.96	52
Tomato Bacterial spot	0.98	0.99	0.98	97
Tomato Leaf Mold	0.97	0.97	0.97	67
Tomato YellowLeaf Curl Virus	0.98	0.95	0.96	114
Tomato mosaic virus	0.99	0.99	0.99	75
Tomato healthy	0.98	0.99	0.99	40

TABLE V. PRECISION, RECALL, F1-SCORE AND SUPPORT FOR EACH CLASS OF THE STUDENT MODEL

Class	Precision	Recall	F1 score	Support
Potato Early blight	0.92	0.98	0.95	45
Potato Late blight	0.93	0.90	0.91	71
Potato healthy	0.93	0.87	0.90	30
Rice Bacterialblight	0.91	0.90	0.90	68
Rice Blast	1.00	1.00	1.00	83
Rice Blight	0.79	0.72	0.76	43
Rice Brownspot	0.93	0.98	0.95	63
Rice Tungro	0.98	1.00	0.99	52
Tomato Bacterial spot	0.99	0.90	0.94	97
Tomato Leaf Mold	0.97	0.93	0.95	67
Tomato YellowLeaf Curl Virus	0.86	0.99	0.92	114
Tomato mosaic virus	1.00	0.95	0.97	75
Tomato healthy	1.00	1.00	1.00	40

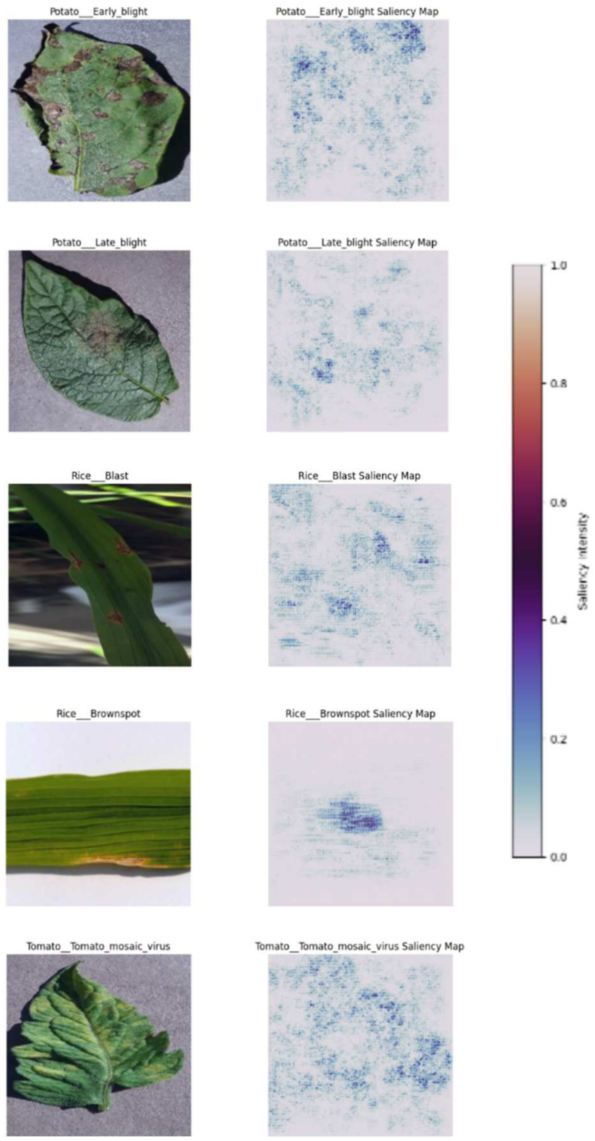

Fig. 9. Examples of DYL-Leaf model's result interpretation with saliency map

TABLE VI. COMPARISON OF PLANT DISEASE CLASSIFICATION MODELS

Authors	Dataset	Purpose	Optimal Model	Accuracy	Trainable Parameters
R. Yakkundimath et al. [11]	Rice leaf (12k augmented)	Rice leaf disease classification	VGG16	92.24%	≈ 138M
H. Yin et al. [12]	Hot pepper leaf (28k)	Hot pepper leaf disease classification	ResNet50	88.38%	≈ 25.6M
Y. Kurmi et al. [13]	PlantVillage (54k)	Bell pepper, potato, tomato leaf disease classification	Custom CNN	92.6-95.8%	≈ 5-10M
S. S. Gaikwad et al. [14]	Fruit leaf (14k)	Fruit leaf disease classification	AlexNet	86.8%	≈ 60M
M. Prabu et al. [15]	Mango leaf (380)	Mango leaf disease classification	MobileNetV2 + SVM	94.5%	≈ 3.5M
This work	PlantVillage (4.1k)	Potato, rice, tomato leaf disease classification	DYL-Leaf (Distilled YOLO-based)	93.8%	545,005

During distillation, the student is guided by the teacher's softened probability distributions (soft labels), which encode rich inter-class relationships. This process compels the student to learn a more compact and generalized representation of the problem space, effectively filtering the teacher's knowledge down to its most essential and robust components. Consequently, the student model becomes less sensitive to the noise inherent in the training set and demonstrates superior generalization to unseen data, leading to its enhanced performance on the validation set. Finally, the comparison presented in Table VI situates DYL-Leaf within the broader research landscape. While other models achieve high accuracy, they often rely on architecture with tens of millions of parameters. DYL-Leaf's ability to deliver state-of-the-art accuracy with a sub-one-million parameter model represents a significant step towards creating truly deployable and accessible AI-powered tools for modern agriculture.

V. CONCLUSIONS

This study successfully demonstrates the efficacy of knowledge distillation in developing a lightweight yet high-performance model for plant leaf disease classification. The distilled student model, DYL-Leaf, with only 545,005 parameters, not only matched but surpassed its 2.6M parameter teacher model. Specifically, DYL-Leaf achieved a superior validation accuracy of 93.8% compared to the teacher's 92.9%, along with improved precision (94.00%), recall (93.23%), and F1-score (93.38%). These findings highlight the potential of KD to create computationally efficient architectures suitable for resource-constrained environments without sacrificing accuracy. While the model performed strongly," Rice Blight" remains a challenging class, suggesting an area for future improvement. This work provides compelling evidence for KD as a viable strategy for creating deployable and effective plant disease classification systems.

ACKNOWLEDGMENT

This research was supported by Universiti Malaysia Pahang Al-Sultan Abdullah International Matching Grant, RDU242726/UIC241528.

REFERENCES

[1] L. Zeifman, S. Hertog, V. Kantorova, J. Wilmoth, and P. Gerland, "A world of 8 billion," United Nations, Department of Economic and Social Affairs, Population Division, Tech. Rep., 2022.

[2] M. Van Dijk, T. Morley, M. L. Rau, and Y. Saghai, "A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050," *Nature Food*, vol. 2, no. 7, pp. 494–501, 2021.

[3] S. Savary, L. Willocquet, S. J. Pethybridge, P. Esker, N. McRoberts, and A. Nelson, "The global burden of pathogens and pests on major food crops," *Nature ecology & evolution*, vol. 3, no. 3, pp. 430–439, 2019.

[4] B. M. Joshi and H. Bhavsar, "Plant leaf disease detection and control: A survey," *Journal of Information and Optimization Sciences*, vol. 41, no. 2, pp. 475–487, 2020.

[5] R. A. Jones, "Global plant virus disease pandemics and epidemics," *Plants*, vol. 10, no. 2, p. 233, 2021.

[6] M. R. Foolad, H. L. Merk, and H. Ashrafi, "Genetics, genomics and breeding of late blight and early blight resistance in tomato," *Critical reviews in plant sciences*, vol. 27, no. 2, pp. 75–107, 2008.

[7] D. R. Walters and J. M. Fountaine, "Practical application of induced resistance to plant diseases: an appraisal of effectiveness under field conditions," *The Journal of Agricultural Science*, vol. 147, no. 5, pp. 523–535, 2009.

[8] A. Khakimov, I. Salakhutdinov, A. Omolikov, and S. Utaganov, "Traditional and current-prospective methods of agricultural plant diseases detection: A review," in *IOP Conference Series: Earth and Environmen- tal Science*, vol. 951, no. 1, 2022, p. 012002.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[10] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[11] R. Yakkundimath, G. Saunshi, B. Anami, and S. Palaiah, "Classification of rice diseases using convolutional neural network models," *Journal of The Institution of Engineers (India): Series B*, vol. 103, no. 4, pp. 1047– 1059, 2022.

[12] H. Yin, Y. H. Gu, C.-J. Park, J.-H. Park, and S. J. Yoo, "Transfer learning-based search model for hot pepper diseases and pests," *Agriculture*, vol. 10, no. 10, p. 439, 2020.

[13] Y. Kurmi, P. Saxena, B. S. Kirar, S. Gangwar, V. Chaurasia, and A. Goel, "Deep cnn model for crops' diseases detection using leaf images," *Multidimensional Systems and Signal Processing*, vol. 33, no. 3, pp. 981–1000, 2022.

[14] S. S. Gaikwad, S. S. Rumma, and M. Hangarge, "Fungi affected fruit leaf disease classification using deep CNN architecture," *International Journal of Information Technology*, vol. 14, no. 7, pp. 3815–3824, 2022.

[15] M. Prabu and B. J. Chelliah, "Mango leaf disease identification and classification using a cnn architecture optimized by crossover-based levy flight distribution algorithm," *Neural Computing and Applications*, vol. 34, no. 9, pp. 7311–7324, 2022.

[16] A. Alidev, "Plantvillage dataset," *Kaggle*, 2021. [Online]. Available: https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset.