SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Security-Aware Self-Supervised Adaptive DCNN for Rice Leaf Disease Detection

R. Dhivya[1], N. Shanmugapriya[2]

[1]Research Scholar, Department of Computer science and Engineering, Dhanalakshmi Srinivasan University, Tiruchirappalli, Tamilnadu, India. E- mail: successdhivya11@gmail.com
[2]Professor, Department of Computer Science and Engineering, Dhanalakshmi Srinivasan University, Tiruchirappalli, Tamilnadu, India. E- mail: shanmugapriyavinod@gmail.com

Abstract
Agriculture plays a critical role in global food security, yet crop production remains highly vulnerable to disease outbreaks that significantly reduce yield and crop quality. Deep learning-based models shows high performance under controlled environment; however, their reliability often deteriorates in real world environments due to variations in illumination, background complexity, and environmental noise, making the model vulnerable to the prediction error and adversarial perturbation. To address these challenges, this study proposes a novel Security-Aware Self-Supervised Adaptive Deep Convolutional Neural Network (SA-SSA-DCNN) for robust rice leaf disease detection. The proposed model uses the self-supervised learning to reduce the model's reliance over the labelled data, attention mechanism to localize the features, and adversarial training to resilient against the environmental perturbation and noise. To evaluate the model's performance the benchmarked rice leaf dataset was obtained from Kaggle. The proposed SA-SSA-DCNN achieved 99.25% accuracy and F1 score as 0.993, Consistently outperforms the standard SSL model and randomly initialized baseline models across all evaluation metrics. Furthermore, the model demonstrates strong robustness under adversarial and noisy conditions, significantly reducing attack success rate while maintaining high prediction reliability for real-world agricultural deployment.

Keywords: Rice Leaf Disease Classification, Deep Learning, Self-Supervised Learning, Adversarial training, Attention mechanism, Environmental noise, Adaptive Mechanism, Precision Agriculture.

Introduction

Agriculture is recognized as one of the backbones of any country in the world, and the agricultural revolution is concurrently evolving with the industrial revolution. Crops play a vital role in our survival. Crop diseases are one of the major causes of food insecurity, which remains one of the most serious global challenges of modern humanity. Plant diseases and pests compromise the food production by 20% to 40%, and in some cases, exceeding 50%, which has a negative influence on our lives and remains a major threat to food security[1]. Thus, crop health is very important for food safety as well as for economic stability. Leaves often serve as an indicator of health of the crops. Plant diseases affecting roots, leaves, stems, and fruits decrease the crop quality and quantity too, which lead to global food shortages and increased food insecurity.

As the crop wastage is increasing due to diseases, there is a growing need for timely and accurate identification of disease. In the current scenario, Artificial Intelligence (AI) is increasingly used in all areas to solve complex problems across various domains including agriculture, to solve complex problems. Deep learning models such as CNN have been widely used to identify plant leaf disease[2]. These methods extract the pathological features like texture, colour, and shape from images to train the classifier to accurately classify the images to different classes of disease. DL models are renowned for their ability to handle large and complex datasets as the agriculture dataset is complex the need of adapting to these deep learning models is becoming increasingly essential.

Models often trained in the controlled lab conditions fails to predict in the field environment due to the domain shift[3]. The factors like environmental noise, varying illumination and background complexity often deteriorates the model's performance in realistic worlds. The standard deep learning models also vulnerable to the small input perturbation that reflects in the model's accuracy which clearly indicates the security weakness and also sensitive to environment noise leading to an overfitting or biased towards a class[4]. Broad evaluations of the deep learning models show continued limitations in both natural and adversarial learning conditions forcing to develop a more adaptive and robust model.

Self-supervised learning models significantly reduces the dependency of the model over the class labels[5]. The attention-based models enhance the model's ability to capture the important features and localization of these features[6][7]. Combination of both these models facilitates the disease recognition across the domains. Despite these advantages, the robustness of such models are underexplored and remains vulnerable to adversarial perturbation and environmental noise[8]. In real world agricultural problems, the model must operate reliably under varying illumination, sensor noise and environmental conditions[9]. Therefore, ensuring the robustness and reliability is essential for the deployment of these models in real world fields. Robust optimization techniques, enhance the model's performance against the perturbation and input variations, offer a promising direction towards achieving a good performance under real-world applications[10]. This drives us to develop an integrated robust learning framework capable of maintaining consistent performance under challenging conditions. To address the limitations, this study proposes a robust framework Security-Aware Self-Supervised Adaptive Deep Convolutional Neural Network for accurate prediction of rice diseases. The proposed research employs the self-supervised learning to reduce the model reliance over labelled datasets, attention mechanism to enhance the localization of disease relevant features, and adversarial training to improve its resilience against the noise and perturbations observed in the real world. The evaluation results demonstrate the proposed method outperforms the state-of-the-art methods like conventional deep learning approaches in terms of prediction as well as reliability, offering a practical solution for the deployment of various field condition in agricultural systems.

Remaining sections of the article is organized as follows. Chapter 2, elaborates the related works on plant disease detection, attention mechanism, self-supervised learning, and robust deep learning approaches. Section 3, describes the proposed SA-SSA-DCNN along with the detailed architecture. Section 4, details the experimental results and discussion over the achieved findings. And section 5, concludes the article with conclusion.

Related Works

Muksimova et al. (2025)[11] developed a novel convolutional neural network (CNN) that integrates an advanced attention mechanism with a trademarked MRI Segmentation Block (MSB) in order to classify Alzheimer's disease efficiently. The MSB, especially designed to manage the skull stripping tasks with high efficiency, while the connection wise attention mechanism allows the network to focus specifically on the crucial regions of brain images associated with different stages of diseases. The proposed architecture achieved high diagnostic accuracy and computational efficiency in terms of complexity makes it to highly suitable for clinical applications where resources are limited.

While moving from laboratory to field conditions, the plant disease classification models often have decreased performance in predicting. A robust framework named Unsupervised Domain Adaptation was proposed to overcome this vulnerability by Tunio et al. (2024)[12]. The local feature extraction strengths of CNNs are combined with global context captured by Mobile Vision Transformers (MViTs). The adversarial learning is employed, hence the model learns domain-invariant representation to minimize the distribution discrepancies, which significantly improves the performance of the classification model on field environment datasets.

In order to protect the deep image classification models from the unknown attacks chen et al. (2025)[13] developed a defense framework, which is a two-stage adversarial purification-based method called latent Noisy Image Construction and Reconstruction (l-NICE). To neutralize perturbation in a unified latent space, the model employs the patch-wise principal component analysis, then the Swin Transformer-based reconstruction network is applied to restore the purified images. This unsupervised approach effectively balances the adversarial robustness and standard accuracy of the model, highlighting the resilience against the grey-box and white-box attacks.

To maximize the accuracy of plant leaf disease, Ali et al. (2024)[14] explored an ensemble of multiple deep learning architectures, which includes DenseNet201 and EfficientNet variants. A novel image processing technique was introduced, which combines the Contrast Limited Adaptive Histogram Equalization (CLAHE) and Adaptive Median Filtering (AMF), which enhances the feature extraction from challenging datasets like PlantVillage. The model achieved near-perfect accuracy 99.89%, which clearly demonstrates how the architectural diversity and pre-processing can create a more stable and diagnostic system.

For plant disease classification Echim et al. (2024)[15] created a model which focused on the intersection point of knowledge distillation, adversarial training and model explainability. The author used the saliency map techniques like Grad-CAM to obtain the insights into how the adversarial attacks diverts the focus away from significant pathological features which reflects in performance of the model. Also, the authors identified that the knowledge distillation can produce student models which are 15-25 times computationally efficient than the teacher models while slightly suffering in the classification performance.

Wang et al. (2024)[16] constructed a dataset called General Plant Infection Dataset (GPID-22) which originally developed to address the limitations of data quantity and diversity considered as one of the largest diversified dataset in this field. They proposed a network architecture (CRE) which integrates the Masked Image Modeling (MIM) with Contrastive Learning. The model learns more generalized image features in an unsupervised manner by utilizing the quantified semantic tokens. These feature facilitates the model having a superior transferability and achieved a state-of-the-art result across 199 different classed in downstream tasks.
To improve the efficiency of semi-supervised adversarial training (SSAT) Ghosh et al. (2025)[17] proposed a data reduction strategy which based on latent clustering. To identify the critical subset of boundary-adjacent data points the method employed k-means clustering which is more important for robust learning. These adjustments make the model to achieve robust accuracies while compared with the models trained on full datasets while using 5 to 10 times less labelled data and significant reduction in the total runtime.
Cagatan et al. (2025)[18] compared the adversarial robustness of seven discriminative self-supervised learning (SSL) methods across different computer vision tasks. Their analysis proved that the SSL model consistently outperformed all other models on ImageNet classification in terms of adversarial robustness, despite the gap narrows during the task specific fine tuning.
Hierarchical Self-Supervised Adversarial Training (HSAT) was introduced by Malik et al. (2024) to address the weakness of vision models in medical imaging. To craft the multi-level adversarial examples HSAT explicitly leverages the patient-slide-patch hierarchy, which are then employed to enforce the robust feature learning across all data under different scales. The results achieved set a new benchmark for microscopic cancer datasets, significantly outperforms the standard instance level adversarial training methods adapted from the natural image domain.
Masked Autoencoder (MAE) is combined with Convolutional Block Attention Module (CBAM) a self-supervised method proposed by Wang et al. (2024)[19] to predict the leaf disease. To learn the robust features from unlabeled data, the model was equipped with MAE, and to capture the sequence relationships between diseased image blocks, a Gated Recurrent Unit (GRU) was utilized. This combination improved the model's sensitivity over pathological features and its ability to generalize the unknown disease types in complex environments.
The limitations identified from the related works are that the existing adversarial paradigms and the SSL significantly lack generalized robustness against unknown attacks, and also maintaining the performance across different domains is still a challenge. To address these limitations Security-Aware Self-Supervised Adaptive Deep Convolutional Neural Network was proposed in the next chapter. The proposed method uses unsupervised representation learning along with security aware adaptive mechanism to have strong resistance to adversarial perturbation and environmental noise, which was further explained in the next chapter.

Methodology

3.1 Proposed SA-SSA-DCNN:

This study proposes a Security-Aware Self-Supervised Adaptive Deep Convolutional Neural Network (SA-SSA-DCNN) designed to achieve the reliable classification performance under the adversarial and clean conditions. The proposed method integrates the self-supervised representation learning as well as adversarial robust optimization to enhance the classification stability and deployment reliability. The architecture follows the two-stage pipeline: 1. Self-Supervised representation learning for enabling the robust learning without the labeled data. 2 Security-Aware adversarial fine-tuning, which makes the model resistance to the adversarial perturbation and real world noise during the supervised learning.
An Adaptive ResNet-18 architecture is employed for the backbone encoder denoted as f(.), Where each convolution layer is augmented using the Convolution Block Attention Module (CBAM). This adaptive attention mechanism enables the model to focus on the disease specific morphological characteristics like discoloration patterns, lesions, and texture variations which excludes the irrelevant background features. This adaptive mechanism improves the representation stability, especially when the disease classes look visually similar or environmental variability.

Adaptive Attention Mechanism:

Plant disease images often contains the variability in background which is caused by some of the factors like illumination change, leaf overlap, or environmental clutter. Conventional image classification models treat all this channel and spatial features equally, which may dilute the disease relevant signals. CBAM address all these limitations by employing the channel attention as well as spatial attentions. The intermediate feature representation is represented in equation 1.

$$\mathbf{F} \in \mathbb{R}^{C \times H \times W} \quad (1)$$

Where C denotes the channel depth, and H, W represents spatial dimensions.

Channel Attention:
The important channel features are identified by the channel attention by aggregating the spatial information through average and max pooling operations followed by a multilayer perceptron is explained in equation 2.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (2)$$

The redefined feature map becomes:

$$F' = M_c(F) \otimes F \quad (3)$$

Equation 3 amplifies the disease relevant channels.

Spatial Attention:
Spatial attention focusing on the important spatial regions as stated in equation 4:

$$M_c(F') = \sigma\left(f^{7X7}([AvgPool(F'); MaxPool(F')])\right) \quad (4)$$

Where f^{7X7} represents a convolution with 7 X 7 kernels.
This mechanism enables the model to accurately localize the infected regions, which improves the classification accuracy across various images that visually similar in look.

3.2 Self-Supervised Representation Learning using SimCLR:

Manual annotation of agricultural datasets are time consuming one and expensive too[20]. To overcome this data dependency limitations, the SimCLR [9] self-supervised learning framework is employed to learn invariant visual representations without requiring labels.

Positive Pair Generation:
For each input image x, two stochastic augmentations $t, t' \sim T$ are applied to produce correlated views $(\tilde{x}_t, \tilde{x}_j)$. These augmentations include random cropping, color jittering, flipping, and Gaussian blur, enabling invariance to environmental variation.

Projection Head Mapping:
Encoder outputs h are mapped into latent space using a non-linear projection head presented in equation 5:

$$z_i = g(h_i) \quad (5)$$

Which improves contrastive representation learning.

Contrastive Objective Function:
The normalized temperature-scaled cross-entropy loss is defined in equation 6:

$$l_{i,j} = -\log \frac{\exp(sim(z_i, z_j)/T)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp(sim(z_i, z_j)/T)} \quad (6)$$

Where cosine similarity is as follows in equation 7:

$$sim(u, v) = \frac{u^T v}{||u||||v||} \quad (7)$$

This learning encourages invariant semantic representations suitable for downstream classification.

3.3 Security Awareness through Adversarial Robustness:

The machine learning or deep learning models trained using the clean data often might fail when subjected to small perturbation which caused by the sensor noise or adversarial manipulations. To address this vulnerability, adversarial training is employed using the Projected Gradient Descent (PGD) [4].

PGD Attack Formulation:
PGD generates adversarial examples iteratively using equation 8 as follows:

$$\mathbf{x_{t+1}} = \prod_{\mathbf{x+s}}(x_t + \alpha . sign(\nabla_{xt} L(\theta, x_t, y)) \quad (8)$$

Where projection $\prod$ ensures perturbation remains within an ϵ-bounded region.

Robust optimization objective:
Training optimizes the following min-max objective stated in equation 9:

$$\min_{\theta} \mathbb{E}_{(x,y)} \max_{||\delta||_\infty \leq \varepsilon} L(\theta, x + \delta, y) \quad (9)$$

The equation 8 forcing the model to learn stable decision boundaries resistant to noise and adversarial perturbations.

3.4 Training Configuration and Hyper Parameter Settings:

Self-Supervised pre-training equips the Adam optimizer with the learning rate of $1 X 10^{-3}$ and cosine annealing schedule. Batch size is set to 256, and training proceeds for 200 epochs using stochastic data augmentations. During the adversarial fine-tuning, learning rate is reduced to $1 X 10^{-4}$, and training continues for 80 epochs with early stopping. PGD parameters include perturbation bound $\epsilon = 8/255$, step size $\alpha = 2/255$ and 10 attack per iteration. Dropout and weight decay ($1 X 10^{-4}$) are used for regularization. A stratified split ensures class balance across the dataset.

3.5 Computational Complexity Analysis:

The primary complexity arises from convolution operations as follows in the equation 10.

$O(N. C. H. W. K^2)$ (10)

Where N is the batch size and K is convolution kernel size.

CBAM introduces negligible overhead due to lightweight pooling operations. Adversarial training increases cost proportionally to PGD iterations but does not affect inference complexity. Thus, deployment efficiency remains suitable for real-time agricultural monitoring.

3.6 Algorithm for proposed SA-SSA-DCNN

Input: Dataset D, encoder f(.), projection head g(.)

Output: Robust disease classification model

1. Initialize encoder parameters.
2. Self-Supervised Pre-training Phase

   For each training batch:

   - Generate two augmented views of each image
   - Extract representations using encoder.
   - Map features through projection head.
   - Compute NT-Xent contrastive loss.
   - Update model parameters via gradient optimization.
3. Remove projection head after convergence.
4. Attach classification layer.
5. Adversarial Fine-Tuning Phase
   - During each training iteration, adversarial samples are generated using the PGD attack.
   - Then it is Fed to the classifier with generated adversarial samples.
   - Model parameters updated through backpropagation by computed entropy loss.
6. To monitor convergence and overfitting the model is evaluated using the validation dataset.
7. Optimized SA-SSA-DCNN model is retained as the final classifier for deployment.

Figure 1 visualize the overall framework of the proposed Security-Aware Self-Supervised Adaptive Deep Convolution Neural Network.

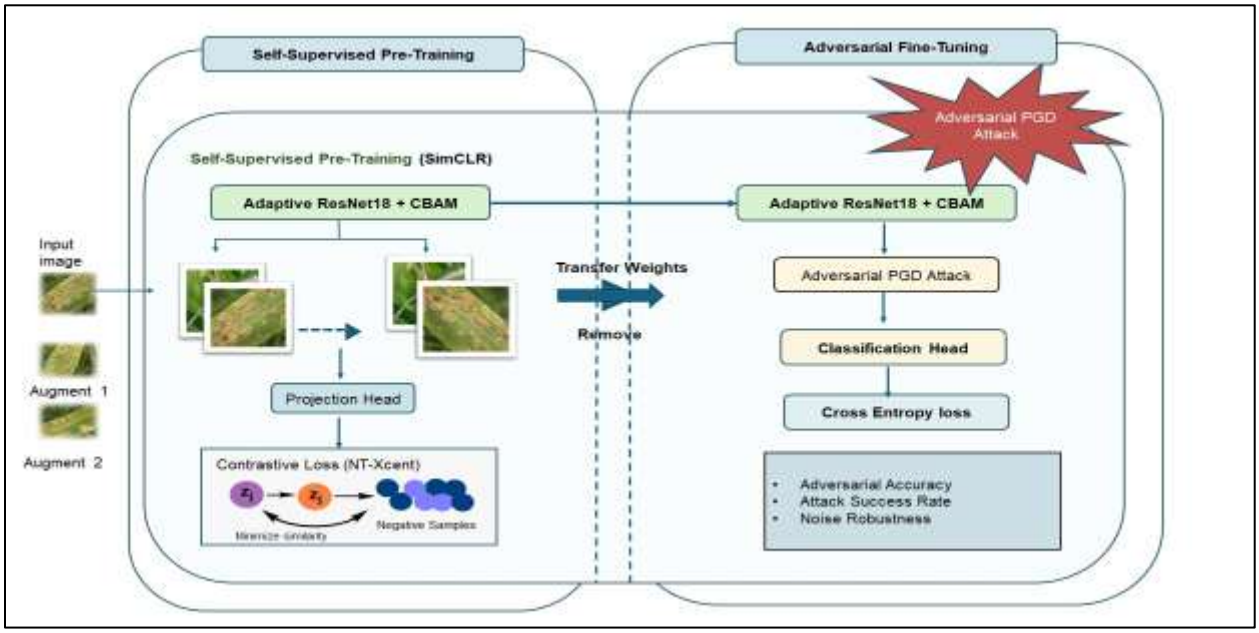

Figure 1: Proposed SA-SSA-DCNN Architecture

4. Results and Discussion

Dataset Description:

For the experiment we have utilized the publicly available benchmarked dataset "Rice Leaf Disease Dataset" from Kaggle repository to evaluate our proposed methodology. The dataset comprises 5932 images which presents the comprehensive array of visuals of major rice leaf diseases providing a suitable platform for training

and evaluating the deep learning models. Key features of the dataset includes, it is relatively well balanced dataset and it has multiple class which represents the pathogenic conditions commonly occur in rice leaves. Images in the dataset representing 4 major rice leaf diseases namely Blast, Bacterial blight, Tungro, Brown Spot and making it a valuable resource for the precision agriculture. Each image in the dataset is annotated with corresponding class label which enables the machine learning models to classify the disease automatically.
Bacterial blight class denotes the images that are affected by bacterial infection which causes wilting and yellowing leaves. Blast class denotes that the rice leaves or affected by fungus which is marked by lesions and necrotic spots on leaves. Brown spot class was characterized by the brown lesions which is often caused by fungal pathogens. Tungro class is a viral disease caused by leafhoppers which leads to discoloration and stunted growth.
Each image in the dataset is provided with the standard color format like RGB which clearly depicting the rice leaves with clear symptoms of disease. The images inside the dataset are typically stored in the standard formats like .jpeg and which may vary in resolution as well as background conditions due the different acquisition devices and field conditions. In the related works often, the images are resized to the fixed dimension of 224 x 224 pixels and the same is carried out in the experiment during preprocessing to reduce the computational complexity. Figure 2 shows the sample images from each class of the dataset and figure 3 depicts the preprocessed images.

Figure 2: Sample images for each class from the dataset

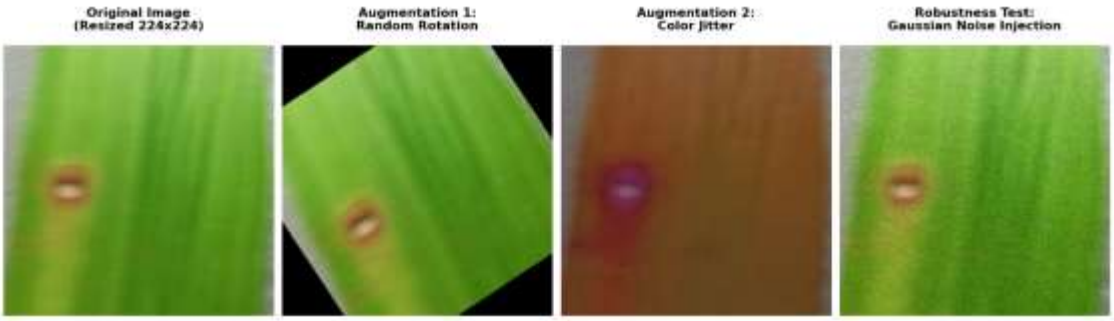

Figure 3: Preprocessed images

Comparison of Standards Classification Metrics:
The proposed model SA-SSA-DCNN is compared with the standard baseline models SSL Standard (ResNet + CBAM) pretrained with SimCLr, but fine-tuned without adversial training and Random Init (ResNet + CBAM) initialized with random weights (no pre-traiing) and trained from scratch. The comparison of standard classification metrics among this models are given in table 1 and the confusion matrix achieved by the models are visualized in figure 4. The proposed SA-SSA-DCNN model shows the strong diagonal dominance which reflects its high accuracy achieved 99.25 %, outperforms the SSL baseline 98.12% and the random initialized model 96.04 %. The prediction demonstrates that the effectiveness of incorporating security-aware adaptation with the self-supervised feature learning, enables more stable representation. A similar pattern also repeated in the F1 score also, the proposed model achieves 0.993 and the baseline models achieved 0.981 and 0.965 respectively which shows how the model balances the precision and recall among the diversified class

distribution. The log loss results further highlight the proposed framework's robustness. The log loss for the proposed model is 0.015 significantly lower, which strongly claims the model's well calibrated probabilistic prediction and its reduction over the uncertainty and overconfidence. While the baseline models SSL standard achieve 0.042 and the random initialization achieved 0.089 which indicates the less reliable confidence estimation. Receiver Operating Characteristic analysis reveals that all the models achieved a high AUC value. Despite the achievement by all the models the proposed model achieved 0.999 which confirms it ability to separate the disease classes. This further enforced by the Mathew Correlation Coefficient, where the proposed model achieved 0.990 and the baseline models achieved 0.975 and 0.952 respectively. Based on the results it is clearly understood that the proposed model is well enough to correlate between the predicted and ground-truth labels. These results clearly indicating that the proposed SA-SSA-DCNN consistently outperforms the baseline models across all the metrics, and validates the benefits of integrating the self-supervised representation learning along with security-aware adaptive mechanism for reliable prediction.

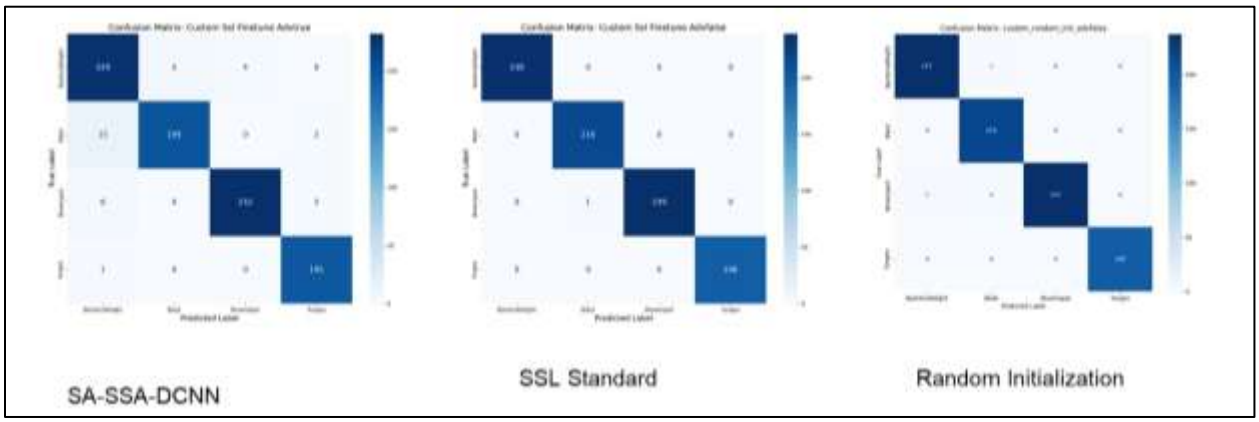

Figure 4. Comparison of confusion matrix for the models used.

Table 1: Validation Metrics

Metric	SA-SSA-DCNN (Proposed)	SSL Standard	Random Init
Accuracy	99.25	98.12	96.04
F1 Score	0.993	0.981	0.965
Log Loss	0.015	0.042	0.089
AUC	0.999	0.998	0.992
MCC	0.990	0.975	0.952

Validation Strategy and Generalization Stability:

To ensure the statistical reliability while maintaining computation efficiency, a stratified hold-out validation was adopted. The data are split into 70% for training, 15% for validation and 15% testing. The stratification scheme ensures that the original class distribution over all the subsets, which ensures the representative sampling of each disease categories. The performance gap between the validation sets and test sets remained under 0.5% irrespective of the all-major metrics, shows the model's stable convergence and the absence of split-specific overfitting. This suggests that the learned representations generalize all the classes effectively beyond the training distribution. The stratified hold out validation provides a practical approximation of generalization stability, especially well suitable for the models trained on large scale image datasets over the k-fold cross validation. The robustness of the models is reflected in low inter-class variance reported in the per class classification analysis which is stated below confirms that the models prediction is not biased towards a specific disease category.

Per-Class Classification:

The per-class classification results stated in table 2 clearly indicating that the models effectiveness over predicting an individual class. The importance of per-class classification is that a model can hide its weakness over individual class prediction still it achieved a higher accuracy rate.

Table 2: Validation metrics per class

Disease Class	SA-SSA-DCNN Accuracy	Standard Accuracy	Improvement
Bacterial Blight	99.4%	98.2%	+1.2%
Blast	98.9%	97.5%	+1.4%
Brown Spot	99.2%	98.4%	+0.8%
Tungro	99.6%	98.8%	+0.8%

The proposed SA-SSA-DCNN achieved 98.9 % across all disease categories clearly demonstrates consistently high class wise accuracy and its capability even among visually similar leaf pathologies. The important improvement observed that is for Blast disease exhibits high inter-class similarity, which highlights the effectiveness of the security aware self-supervised representation. The models show relatively uniform performance over all the classes suggested that the model does not dominant over specialize towards a particular disease. While the dataset contain only diseased samples, it shows the low-off diagonal confusion in the confusion matrix indicates the effective separation among the disease categories which indirectly highlights the high classification reliability.

Security and Robustness metrics:

This section elaborates how well the model performed under the hostile and real world noisy environment not just on the clean test data. The proposed model and the baseline models compared in terms of Attack Success Rate (percentage of adversarial attacks that successfully fooled the model), Adversarial Accuracy (images that are intentionally modified using a PGD adversarial attack), and Noise Robustness (images are corrupted with Gaussian noise simulating real-world disturbance). Proposed model achieved 3.13 as ASR whereas the baseline models achieved 42.5% and 65.4% respectively, that clearly denotes the model is secured than the other models. As far as the adversarial accuracy is concerned, the proposed model achieved 96.88% and the baseline models achieved 57.5% and 34.6 % respectively supporting that even the images are getting attacked the proposed model predicts well. While reviewing the noise robustness the proposed model achieved 98.05% where the SSL achieved 55.5% and the Random initialization model achieved 32.2% clearly indicating that the model works reliably on the real agricultural field conditions.

Self-Supervised Pre-training vs Random Initialization:

This section evaluates how the different initialization affects the learning of the models used in this experiment. The proposed model achieved 99.25%, standard SSL method achieved 98.12% and the random initialization model 96.04% clearly indicates the model trained with self-supervised pre-training start with best internal feature representations than the models which starts with random weights. Robustness comparison achieved by the models clearly indicates that the random initialization model learns the superficial patterns, standard SSL model learns the generalized features whereas the proposed model optimizes both the features. The saliency analysis stated in figure 5, confirms that the proposed model predominantly focusing on the infected leaf regions but the other models often attends to the broader leaf areas or background, which makes the proposed model to make the biologically meaningful decision-making and improves the trustworthiness while compared with other models used in this experiment.

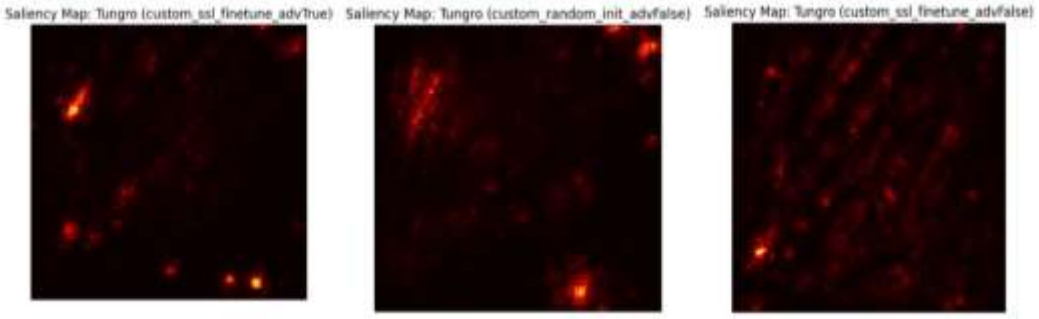

Figure 5. Saliency Analysis for the models used in this research

The experimental findings clearly demonstrate that the proposed SA-SSA-DCNN framework not only introduce the improvements in prediction beyond the conventional accuracy. Unlike the traditional CNN models that relies on supervised learning or random initialization the model integrates the self-supervised representation learning along with security aware adaptive optimization enables the extraction of stable and discriminative

disease features and allow the model to have the strong resistance over adversarial perturbation and environmental noise. Furthermore, the saliency analysis confirms the model focuses on biologically meaningful regions over the background makes the model to take the improved decision interpretability and reliability. The high class accuracy, robust adversarial resistance, balance per-class sensitivity, and the improved prediction calibration establishes the model is more reliable and making it suitable for practical deployment in real world agricultural monitoring system.

Conclusion

This study presented a robust deep learning framework for accurately predict the rice lead disease which focuses on precision agriculture. By adapting advanced features of deep learning along with attention mechanism and adversarial training the model effectively captured the disease relevant patterns across disease categories. Experimental evaluation conducted on a benchmark rice disease dataset obtained from Kaggle and the model outperformed all the baseline models used in this research. The results prove that the proposed framework can significantly contribute to automated crop monitoring system enables the timely identification of disease accurately. Furthermore, promising results under challenging conditions shows that the capability of the model to withstand environment perturbation and sensor noise ensuring reliable performance in real-world deployment scenario.

References

[1] Y. Kaya and E. Gürsoy, "A review of deep learning architectures for plant disease detection," Turkish J. Biol., vol. 49, no. SI-5, pp. 459–497, 2025, doi: 10.55730/1300-0152.2761.

[2] S. Mohanty, D. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," Front. Plant Sci., 2016.

[3] M. Shoaib, A. Sadeghi-Niaraki, F. Ali, I. Hussain, and S. Khalid, "Leveraging deep learning for plant disease and pest detection: a comprehensive review and future directions," Front. Plant Sci., vol. 16, no. February, pp. 1–19, 2025, doi: 10.3389/fpls.2025.1538163.

[4] Z. Luo, Q. Li, and J. Zheng, "A study of adversarial attacks and detection on deep learning-based plant disease identification," Appl. Sci., vol. 11, no. 4, pp. 1–16, 2021, doi: 10.3390/app11041878.

[5] X. Huan, B. Chen, and H. Zhou, "A Unified Self-Supervised Framework for Plant Disease Detection on Laboratory and In-Field Images," Electron., vol. 14, no. 17, Sep. 2025, doi: 10.3390/electronics14173410.

[6] P. Alirezazadeh, M. Schirrmann, and F. Stolzenburg, "Improving Deep Learning-based Plant Disease Classification with Attention Mechanism," Gesunde Pflanz., vol. 75, no. 1, pp. 49–59, 2023, doi: 10.1007/s10343-022-00796-y.

[7] X. Ma, W. Chen, and Y. Xu, "ERCP-Net: a channel extension residual structure and adaptive channel attention mechanism for plant leaf disease classification network," Sci. Rep., vol. 14, no. 1, pp. 1–14, 2024, doi: 10.1038/s41598-024-54287-3.

[8] S. Karthikeyan, R. Charan, S. Narayanan, and L. Jani Anbarasi, "Enhanced plant disease classification with attention-based convolutional neural network using squeeze and excitation mechanism," Front. Artif. Intell., vol. 8, no. August, 2025, doi: 10.3389/frai.2025.1640549.

[9] X. Jiang et al., "PlantCaFo: An efficient few-shot plant disease recognition method based on foundation models," Plant Phenomics, vol. 7, no. 1, Mar. 2025, doi: 10.1016/j.plaphe.2025.100024.

[10] H. Shadab Malik, "Hierarchical Self-Supervised Adversarial Training for Robust Vision Models in Histopathology." [Online]. Available: https://github.com/HashmatShadab/HSAT.

[11] S. Muksimova, S. Umirzakova, N. Iskhakova, A. Khaitov, and Y. I. Cho, "Advanced convolutional neural network with attention mechanism for Alzheimer's disease classification using MRI," Comput. Biol. Med., vol. 190, May 2025, doi: 10.1016/j.compbiomed.2025.110095.

[12] S. Mohanty et al., "Self-supervised contrastive learning on agricultural images," Comput. Electron. Agric., vol. 14, no. 1, p. 107967, Jun. 2025, doi: 10.1016/j.compag.2021.106510.

[13] Z. Duan, S. Xu, B. Chen, C. Wang, and M. Zhou, "TopicNet: Semantic graph-guided topic discovery," arXiv, 2021.

[14] A. H. Ali, A. Youssef, M. Abdelal, and M. A. Raja, "An ensemble of deep learning architectures for accurate plant disease classification," Ecol. Inform., vol. 81, Jul. 2024, doi: 10.1016/j.ecoinf.2024.102618.

[15] S.-V. Echim, I.-M. Tăiatu, D.-C. Cercel, and F. Pop, "Explainability-Driven Leaf Disease Classification Using Adversarial Training and Knowledge Distillation," Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.00334

[16] Z. Wang, R. Wang, M. Wang, and M. Zhang, "method with General Plant Infection dataset," vol. 4.
[17] S. Ghosh, Y. Xu, and X. Zhang, "Efficient Semi-Supervised Adversarial Training via Latent Clustering-Based Data Reduction," Oct. 2025, [Online]. Available: http://arxiv.org/abs/2501.10466
[18] Ö. V. Çağatan, Ö. F. Tal, and M. E. Gürsoy, "Adversarial Robustness of Discriminative Self-Supervised Learning in Vision," Sep. 2025, [Online]. Available: http://arxiv.org/abs/2503.06361
[19] Y. Wang et al., "Classification of Plant Leaf Disease Recognition Based on Self-Supervised Learning," Agronomy, vol. 14, no. 3, Mar. 2024, doi: 10.3390/agronomy14030500.
[20] M. Wu et al., "SinGAN-CBAM: a multi-scale GAN with attention for few-shot plant disease image generation," Front. Plant Sci., vol. 16, no. November, pp. 1–20, 2025, doi: 10.3389/fpls.2025.1703529.