



An Explainable Dual-Stream Center-Weighted Smart Zoom Architecture for Robust Potato Leaf Disease Classification

Hitendra Singh^{1*}, Ajay Kumar Sharma² and Mayank Patel³

^{1*}Department of CSE, GITS, Udaipur, 313003, Rajasthan, India.

²Department of CSE, GITS, Udaipur, 313003, Rajasthan, India.

³Department of CSE, GITS, Udaipur, 313003, Rajasthan, India.

*Corresponding author(s). E-mail(s): shitendra44@gmail.com;

Contributing authors: profsharmaak@gmail.com;

mayank999_udaipur@yahoo.com;

Abstract

Accurate identification of potato leaf diseases is very important for reducing potato crop loss. Deep learning models get excellent accuracy on lab controlled datasets but they do not achieve the same accuracy on real-world datasets, because they have complex background. Most existing methods use information from the entire leaf and try to predict the result. But they do not pay sufficient attention to the diseased portion of the leaf.

This paper presents a **Dual-Stream Center-Weighted Smart Zoom Architecture** for strong potato leaf disease classification. This architecture combines a Center-Weighted Multi-Scale Attention Module with a Gaussian center prior to highlight disease-affected regions and ignore background. This attention map guides to a weakly supervised Smart Zoom mechanism that automatically extracts disease related region (Region of interest) from the full image. This extracted ROI is analyzed with the original image using a shared EfficientNetV2B0 backbone in a dual-stream architecture. A SHA-256-based dataset preparation mechanism removes duplicate images and verifies that there is no data leakage in training, validation, and testing subsets.

This proposed architecture was tested on the PlantVillage, PLD.Pakistan, and PLD Indonesia datasets. On the challenging PLD Indonesia dataset, it achieved an overall accuracy **88.12%**, a balanced accuracy **88.25%**, and an MCC **0.8555**. For model reliability, Grad-CAM visualizations are integrated in this architecture, so that we can visualize that model focuses on diseased spot or not.

Overall, the proposed architecture provides an accurate and reliable solution for automated potato leaf disease classification.

Keywords: Potato Disease Detection, Deep Learning, EfficientNetV2B0, Attention Mechanism, Smart Zoom, Explainable AI, Cross-Dataset Generalization.

1 Introduction

Potato (*Solanum tuberosum*) is one of the most cultivated food crops in the world and it is very important for global food security because it has high nutritional value and it can grow in diverse climate conditions. According to the Food and Agriculture Organization (FAO), Potato cultivation is a very important source of income for millions of farmers in the world. So that, accurate potato disease detection on time is very important for improving potato production. [1, 2].

Recent advancements in deep learning have improved automated plant disease classification. Convolutional Neural Networks (CNNs) and transfer learning models, like ResNet, DenseNet, EfficientNet, and EfficientNetV2, have achieved high accuracy in disease classification by learning from visual representations of leaf images [3–5]. Recently, Attention mechanisms, Vision Transformers (ViTs), and Explainable Artificial Intelligence (XAI) have further enhanced classification performance and model interpretability by focusing on only disease relevant regions and providing visual explanations for their predictions to increase model reliability [6–8].

However, most studies have been evaluated using laboratory-controlled datasets such as PlantVillage[9], which have uniform backgrounds and controlled lighting conditions. These datasets get very high accuracy but they do not represent real world agricultural environment. Real-world datasets such as PLD_Indonesia contain cluttered backgrounds, lighting variations, shadows, and multiple disease categories, so that it makes disease classification very challenging [10]. Also, existing methods mainly depend on global feature extraction and only result explainability. They do not include lesion localization, region-of-interest (ROI) extraction, duplicate removal and data leakage verification.

To solve these problems, this paper proposes a **Dual-Stream Center-Weighted Smart Zoom Architecture** for strong and explainable potato leaf disease classification. The proposed architecture combines a Center-Weighted Multi-Scale Attention Module with a Gaussian center prior to highlight disease affected regions and ignore background interference. The generated attention map guides to a weakly supervised Smart Zoom mechanism, so that it automatically extracts disease relevant ROIs. Both the ROIs and the original image are analyzed using a shared EfficientNetV2B0 architecture. Also, a SHA-256 based dataset preparation mechanism removes duplicate images from the dataset and it verifies that there is no overlap in between the training, validation, and testing dataset. The architecture provides visualizations of attention map in local (ROI) and global (original) image for transparency.

Existing approaches mainly highlight either attention mechanisms or only results explainability. But the proposed architecture provides lesion guided attention, ROI

extraction, dual-stream feature learning, SHA-256 based dataset cleaning, no data leakage verification, and multi-level explainability within a single architecture.

Also, the model was tested on three publicly available datasets PlantVillage[9], PLD_Pakistan[11], and PLD-Indonesia[10] datasets and achieved high accuracy in all of these. It includes hyperparameter optimization, architectural ablation studies, cross dataset analysis, and explainability. The experimental results demonstrate that the proposed framework provides an accurate, and reliable solution for automated potato leaf disease classification in lab controlled environment and uncontrolled environment(real world).

1.1 Research Contributions

The main contributions of this work are as follows:

1. A **Center-Weighted Multi-Scale Attention Module** combines with a Gaussian center prior to improve lesion localization and ignore background.
2. A weakly supervised **Smart Zoom mechanism** that automatically extracts disease relevant area(ROIs).
3. A shared-weight **Dual-Stream EfficientNetV2B0** architecture that learns from global information and local information simultaneously.
4. A **SHA-256 based dataset cleaning mechanism** for duplicate image removal to improve the reliability.
5. An **Explainable AI framework** which combining Attention Maps, Global Grad-CAM, and Local Grad-CAM for transparent disease classification.

2 Related Work

2.1 Deep Learning-Based Plant Disease Detection

Deep learning has become a main approach for automated plant disease classification by using convolutional neural networks (CNNs) to learn directly from leaf images. Architectures such as AlexNet, VGG16, ResNet, and DenseNet etc. are conventional machine learning methods. Transfer learning improves performance by using ImageNet pre-trained models for plant disease classification.

Modern CNN architectures, EfficientNet and EfficientNetV2 provide an effective balance between classification accuracy and computational efficiency. Mozumder *et al.* [3], Fathonah *et al.* [4], and González-Briones *et al.* [5] demonstrated that EfficientNet-based models provide strong classification performance while maintaining computational efficiency. These methods use global feature extraction and include background for learning.

2.2 Attention-Based Disease Detection

Attention mechanisms have been widely used to improve feature representation in complex field conditions. Nazir *et al.* [12] incorporated attention into EfficientNetV2B4 for potato disease classification, while Kiruthika *et al.* [6], Nasra *et al.* [13], and Wahyunigrum *et al.* [14] used CBAM with Attention mechanism in CNN architectures to

improve the focus on lesion and ignore other background. These approaches improve feature learning by using feature maps within a single-stream architecture.

2.3 Explainable Artificial Intelligence for Plant Disease Diagnosis

Explainable Artificial Intelligence (XAI) is a very important framework for improving the transparency of deep learning-based plant disease classification. Alhammad *et al.* [8], Nurullah *et al.* [7], and Hoqueet *et al.* [15] show that Grad-CAM based visualization easily highlights disease relevant regions and improves the model interpretability and reliability.

Most studies use Grad-CAM for only results visualization after prediction. But the proposed architecture uses it for Center-Weighted Multi-Scale Attention, Smart Zoom based lesion localization, and both Global and Local Grad-CAM visualizations to give multi level explainability in the disease classification process.

2.4 Research Gap

Most studies are conducted on lab controlled datasets and they get very high accuracy. But they are less useful in real field because of background clutter and lighting variation and due to other factors. Only few studies are conducted on real-world datasets. So it is very important to shift research from lab controlled datasets to real-world datasets real field datasets. Shabrina *et al.* [10] introduced the PLD_Indonesia dataset, It contains seven potato disease categories which are collected under uncontrolled field conditions with complex backgrounds, and lighting variations. Also, Sinamenye *et al.* [16] proposed an architecture for this PLD_Indonesia dataset and improved classification performance by combining CNN and Vision transformer. These approaches still depend on global feature extraction without using lesion-guided localization, and ROI extraction.

For these limitations, the proposed architecture combines Center-Weighted Smart Zoom Architecture, a Multi-Scale Attention Module, a weakly supervised Smart Zoom mechanism for ROI extraction, and an Explainable AI framework within a single architecture. Also, the proposed architecture is tested on PlantVillage, PLD_Pakistan, and PLD_Indonesia datasets and performs very well on controlled, semi-controlled, and real-world environments.

3 Materials and Dataset Preparation

The proposed architecture uses three publicly available potato leaf disease datasets. Each dataset represents different environmental complexity.

3.1 Datasets

3.1.1 PlantVillage Dataset

The PlantVillage dataset has controlled laboratory conditions. It contains 2,152 images divided into three classes (Early Blight, Late Blight, and Healthy). All images were

Table 1 Summary of Duplicate Image Removal

Dataset	Original Images	Duplicates Removed	Final Images
PlantVillage	2152	0	2152
PLD_Pakistan	4072	16	4056
PLD_Indonesia	3076	55	3020

captured under uniform background and lighting. This dataset is widely used for plant disease classification but it has controlled environmental conditions so it does not present practical agricultural conditions[9].

3.1.2 PLD_Pakistan Dataset

The PLD_Pakistan dataset has a semi-controlled environment between laboratory and field conditions. It contains 4,072 images of Early Blight, Late Blight, and Healthy leaves which were collected from potato fields in Punjab, Pakistan[11].

3.1.3 PLD_Indonesia Dataset

The PLD_Indonesia dataset was selected as the primary dataset because it has real agricultural environments. It contains 3,076 images belonging to seven disease categories and it has complex backgrounds, illumination variations, shadows, overlapping leaves, and other field challenges [10].

3.2 Dataset Preparation

Duplicate images were removed before dataset partitioning using the Secure Hash Algorithm (SHA-256). As shown in Table 1, no duplicates were found in PlantVillage, 16 duplicate images were removed from the PLD_Pakistan and 55 duplicate images were removed from the PLD_Indonesia datasets.

After that, each dataset was divided into three subsets training (80%), validation (10%), and testing (10%). Zero overlap in these subsets was verified through SHA-256 to prevent data leakage.

All images were resized to 224×224 pixels and normalized using the EfficientNetV2 `preprocess_input` function. Random horizontal flipping and random vertical flipping were applied only on the training subset. Also, class imbalance was solved using inverse frequency class weights in the sparse categorical cross-entropy loss function.

4 Proposed Dual-Stream Smart Zoom Architecture

4.1 Overall Framework

Figure 1 shows the proposed Dual-Stream Center-Weighted Smart Zoom Architecture. The input is a potato leaf image of size $224 \times 224 \times 3$, a shared EfficientNetV2B0 backbone extracts high-level features. These features are processed by the Center-Weighted Multi-Scale Attention Module to generate an attention map and highlights

disease relevant regions while ignoring background. Global Average Pooling (GAP) converts the feature maps into a global feature vector ($1 \times D$).

The attention map guides the Smart Zoom Module to identify the disease relevant lesion region (ROIs) and extract it. The ROI is processed by the EfficientNetV2B0 backbone. After that, Global Average Pooling is used to get a local feature vector ($1 \times D$).

The global feature vectors and local feature vectors are concatenated. The features were first normalized using Batch Normalization. Then, a Dropout layer (0.4) was used to reduce overfitting. Finally, the features were passed to a fully connected layer with 128 ReLU neurons and L2 regularization, followed by the Softmax layer for final classification.

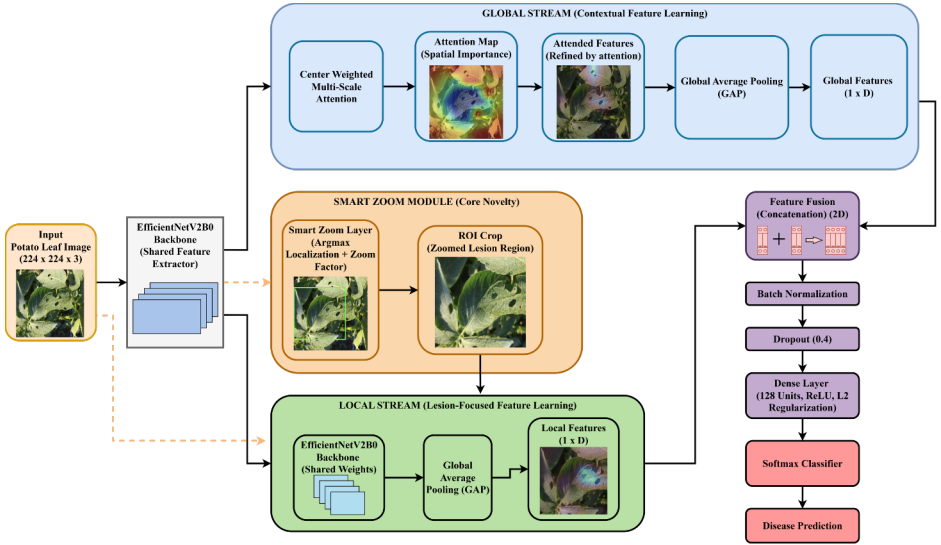


Fig. 1 Overview of the proposed Dual-Stream Center-Weighted Smart Zoom Architecture.

4.2 Global Feature Extraction Branch

The Global Feature Extraction Branch uses the EfficientNetV2B0 backbone to extract high-level features which describe global disease patterns, and contextual information required for reliable classification.

4.3 Center-Weighted Multi-Scale Attention Module

The proposed Center-Weighted Multi-Scale Attention Module reduces the effect of the background and focuses more on the disease-affected areas of the leaf. It uses average pooling and max pooling to extract important spatial information. This information

is then processed using parallel 3×3 and 7×7 convolution layers to capture disease features at different scales. Finally, a Gaussian center prior gives more importance to the central part of the image and generates an attention map for selecting the most relevant region of interest (ROI).

4.4 Smart Zoom Layer

The Smart Zoom Layer automatically identifies the most important disease-affected region without using manually marked bounding boxes. It uses the attention map to locate the area with the most useful disease information and creates a bounding box based on the predefined zoom factor. The selected region is then resized using a crop-and-resize operation.

4.5 Local Feature Extraction Branch

The extracted ROI is passed through this branch and shares the same EfficientNetV2B0 backbone as the Global Branch. Sharing of the backbone weights reduces the number of model parameters and keeps feature extraction consistent in both branches. This branch focuses on disease features such as lesion boundaries, and damaged tissue.

4.6 Dual-Stream Feature Fusion

The features extracted from the Global Branch and the Local Branch are merged together. The combined features are normalized using Batch Normalization and passed through a Dropout layer to improve generalization. They are then processed by a fully connected layer with ReLU activation and L2 regularization. Finally, the Softmax layer predicts the disease class.

4.7 Three-Stage Progressive Fine-Tuning

The ImageNet pre-trained backbone is used for potato leaf disease classification using a three-stage fine-tuning strategy. First, only the attention module, feature fusion, and classification layers are trained while keeping the backbone frozen. After that, the upper layers of the backbone are gradually unfrozen and trained with smaller learning rates. The Batch Normalization layers are kept frozen during all training stages.

4.8 Integrated Explainability Framework

The proposed model includes an Explainable AI (XAI) framework to improve the understanding of predictions. The attention map finds the disease affected region. Global Grad-CAM highlights the relevant areas in the full leaf image, while Local Grad-CAM highlights the relevant areas in the Smart Zoom ROI. These visualizations make the model's prediction more transparent and increase the reliability of the proposed model.

5 Experimental Setup

5.1 Experimental Environment

All experiments were conducted on a PC with Microsoft Windows 11 (64-bit), an 11th Generation Intel® Core™ i5-1135G7 processor, and 16 GB RAM. As no dedicated GPU was available, the model was trained and tested on the CPU. The proposed framework was developed in Python using TensorFlow and Keras. OpenCV was used for image preprocessing. NumPy, Pandas, and Scikit-learn were used for data processing and model evaluation. The hardware and software configuration is given in Table 2.

Table 2 Experimental Environment

Component	Configuration
Operating System	Windows 11 (64-bit)
Processor	Intel Core i5-1135G7 (11th Gen)
CPU	4 Cores / 8 Threads
Memory	16 GB RAM
GPU	Not Available
Python	3.13.5
TensorFlow / Keras	2.20.0 / 3.10.0
OpenCV	4.12.0
NumPy / Pandas	2.4.2 / 2.2.3
Scikit-learn	1.6.1
Base Model	EfficientNetV2B0 (ImageNet)

5.2 Hyperparameter Configuration

The proposed architecture was optimized by many experiments on the PLD-Indonesia dataset. All input images were resized to $224 \times 224 \times 3$ pixels, and a batch size was set to 32 for training. Some previous studies have found that adaptive optimizers perform better than traditional optimizers. For example, Charisma *et al.* [17] compared SGD, RMSprop, and Adam and found that Adam achieved the best classification performance. Based on these findings, the proposed architecture uses the Nadam optimizer, which improves Adam by adding Nesterov accelerated gradients for faster and more stable training. A progressive learning rate schedule, dropout regularization, and a three-stage fine-tuning strategy were used to improve training performance. The tested hyperparameter values and the final selected values are described in Section 6.1.

5.3 Evaluation Metrics

The proposed architecture was evaluated by Accuracy, Precision, Recall, F1-score, Balanced Accuracy, Matthews Correlation Coefficient (MCC), and Cohen's Kappa to

provide a complete assessment of classification performance. These metrics were computed from the confusion matrix, which consists of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Accuracy reflects the overall classification performance. Balanced Accuracy, MCC, and Cohen’s Kappa provide more reliable measures for class-imbalanced datasets. Together, these metrics provide a complete evaluation of the proposed framework on the PlantVillage, PLD_Pakistan, and PLD_Indonesia datasets.

6 Results

The proposed architecture was evaluated by hyperparameter optimization, architectural ablation, cross-dataset generalization, and explainability analysis. The PLD_Indonesia dataset was selected as the main benchmark because it represents real-world conditions.

6.1 Hyperparameter and Architectural Optimization

The architectural design and training hyperparameters were systematically optimized using the PLD_Indonesia dataset. The final configuration adopted in a series of experiments is summarized in Table 3.

Table 3 Summary of Architectural and Training Hyperparameter Optimization

Parameter	Tested Values	Selected
<i>Architectural Optimization</i>		
Gaussian Prior (σ)	0.1, 0.3, 0.5, 0.7, 1.0	0.7
Smart Zoom Factor	0.3, 0.5, 0.7	0.7
Fine-Tuning Depth	15/30, 20/35, 25/40	25/40
<i>Training Optimization</i>		
Dropout Rate	0.3, 0.4, 0.5	0.4
Dense Layer Units	64, 128, 256	128
Learning Rate Schedule	Three schedules	$1 \times 10^{-3} \rightarrow 3 \times 10^{-4} \rightarrow 2 \times 10^{-4}$
Optimizer	Adam, Nadam	Nadam

Notes: (i) Fine-Tuning Depth “15/30”, “20/35”, and “25/40” denote the number of backbone layers unfrozen during Stages 2 and 3, respectively. (ii) The selected learning rate schedule corresponds to the three-stage progressive fine-tuning strategy. (iii) σ denotes the standard deviation of the Gaussian center prior used in the Center-Weighted Multi-Scale Attention Module. (iv) The selected values were consistently adopted in all subsequent experiments.

6.2 Architectural Ablation Study

To understand the effect of each component, an ablation study was performed on the proposed architecture. As shown in Table 4, the Center-Weighted Multi-Scale Attention Module improved the classification accuracy from 83.17% to 84.82%. The

Smart Zoom mechanism provided an additional improvement, increasing the accuracy to 85.48%. The best results were obtained when all the proposed components were combined in the Dual-Stream architecture, which achieved an accuracy of 88.12% and a balanced accuracy of 88.25%. Compared with the baseline model, the complete framework improved the classification accuracy by 4.95 percentage points.

Table 4 Architectural Component Ablation (On PLD_Indonesia Dataset)

Experiment ID	A1	A2	A3	A4
Model Configuration	Baseline	+ MSA	+ Smart Zoom	+ DSF (Proposed)
Accuracy (%)	83.17	84.82	85.48	88.12
Balanced Accuracy (%)	82.64	85.00	85.80	88.25
Macro Average				
Precision	0.8245	0.8419	0.8504	0.8765
Recall	0.8264	0.8500	0.8580	0.8825
F1 Score	0.8206	0.8413	0.8527	0.8751
Cohen Kappa Score	0.7939	0.8145	0.8228	0.8545
Matthews Correlation Coefficient	0.7949	0.8160	0.8231	0.8555

Note: MSA = Multi-Scale Attention; DSF = Dual Stream Fusion.

6.3 Final Model Performance

The proposed framework achieved 88.12% accuracy, 88.25% balanced accuracy, a Cohen’s Kappa score of 0.8545, and an MCC of 0.8555 on the PLD_Indonesia dataset. As shown in Fig. 2, the training and validation curves remained stable throughout the three-stage fine-tuning process, with no clear signs of overfitting.

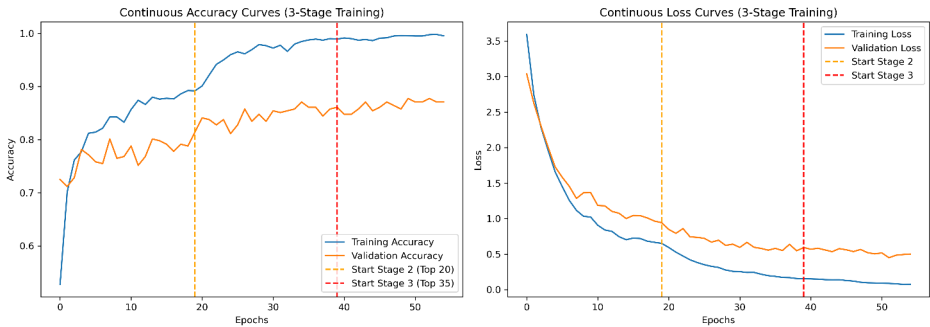


Fig. 2 Training and validation accuracy/loss curves during the three-stage progressive fine-tuning process on the PLD_Indonesia dataset. The vertical dashed lines indicate the transition between successive fine-tuning stages.

6.4 Confusion Matrix Analysis

Figure 3 shows that the model correctly classified most disease classes. Only a few classes were observed incorrectly due to visual similarity. It indicates that the proposed architecture works well on real-world images.

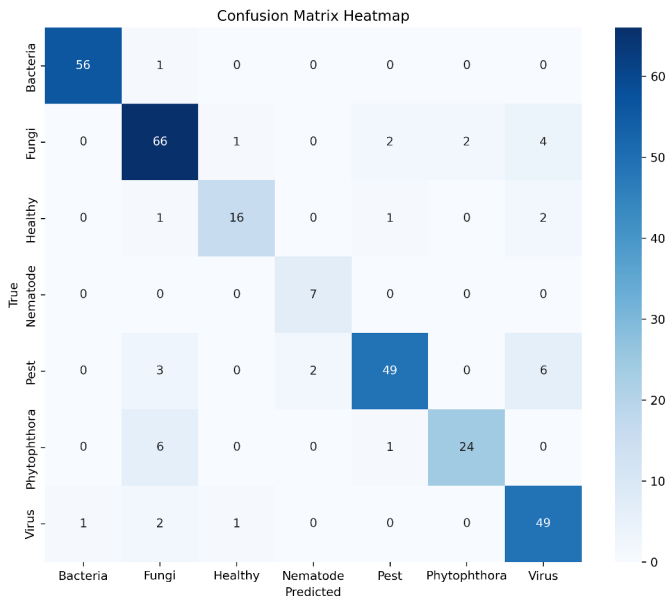


Fig. 3 Confusion matrix of the proposed framework evaluated on the PLD_Indonesia dataset. The majority of samples are correctly classified, indicating strong class-wise discrimination performance.

6.5 Explainability Analysis

The explainability of the proposed architecture was analysed using the attention module along with Global and Local Grad-CAM visualizations. As shown in Fig. 4, the attention module correctly highlights the disease-affected region, while the Smart Zoom layer extracts the most important region of interest (ROI). Both Global and Local Grad-CAM focus mainly on the lesion area instead of the background. This makes it easier to understand how the model makes its predictions and increases the reliability of the proposed framework.

6.6 Cross-Dataset Generalization Analysis

Cross-dataset experiments were conducted on the PlantVillage and PLD_Pakistan datasets to evaluate the reliability of the proposed architecture during domain shift. As shown in Table 5, the architecture achieved 100.00% accuracy on the PlantVillage dataset and 99.51% accuracy on PLD_Pakistan dataset, when it was trained and tested

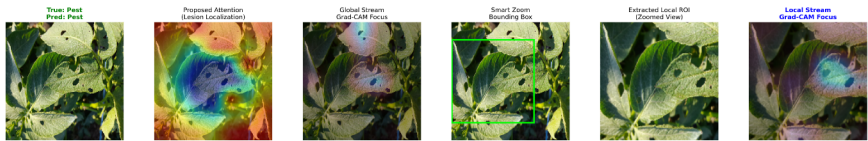


Fig. 4 Representative explainability analysis of the proposed Dual-Stream Center-Weighted Smart Zoom Architecture showing (a) input image, (b) attention map, (c) Global Grad-CAM, (d) Smart Zoom localization, (e) extracted ROI, and (f) Local Grad-CAM visualization.

on the same dataset. When training and testing were performed on different datasets, the accuracy dropped to 55.58% and 60.36%. These results show that domain shift has a major effect on model performance and highlight the need to evaluate plant disease classification models on real-world datasets rather than depending only on laboratory controlled datasets.

The obtained results are consistent with the findings reported by Saha *et al.* [18]. They also confirm that cross-dataset evaluation is much more challenging than evaluation on the same dataset.

Table 5 Model Testing for PlantVillage and PLD_Pakistan Dataset with Cross-Dataset Generalization

Experiment ID	PV	PLDP	CD1	CD2
Training Dataset	PlantVillage	PLD_Pakistan	PlantVillage	PLD_Pakistan
Testing Dataset	PlantVillage	PLD_Pakistan	PLD_Pakistan	PlantVillage
Accuracy (%)	100.00	99.51	55.58	60.36
Balanced Accuracy (%)	100.00	99.59	58.92	71.19
Macro Average				
Precision	1.0000	0.9944	0.6668	0.7828
Recall	1.0000	0.9959	0.5892	0.7119
F1 Score	1.0000	0.9951	0.4974	0.6262
Cohen Kappa Score	1.0000	0.9925	0.3509	0.3032
Matthews Correlation Coefficient	1.0000	0.9925	0.4107	0.4306

6.7 Comparison with Existing Methods

Table 6 compares the performance of the proposed architecture with existing methods evaluated on the PLD_Indonesia dataset. Shabrina *et al.* [10] introduced this dataset and reported an accuracy of 73.60% using EfficientNetV2B3. After that, Sinamenye *et al.* [16] improved the accuracy to 85.06% by combining EfficientNetV2B3 with a Vision Transformer. However, neither study reported duplicate image removal or data leakage verification.

The proposed architecture solves these limitations by using SHA-256-based duplicate removal, data leakage verification, attention-based ROI localization, dual-stream

feature learning, and Explainable AI into a single architecture. As a result, it achieved the highest classification accuracy of **88.12%** on the PLD_Indonesia dataset.

Table 6 Comparison with Representative Methods on the PLD_Indonesia Dataset

Method		Dataset Cleaning	ROI	XAI	Accuracy (%)
Shabrina al. [10]	et	×	×	×	73.60
Sinamenye al. [16]	et	×	×	✓	85.06
Proposed		✓ (SHA-256, 55 duplicates removed)	✓	✓	88.12

Notes: (i) Dataset Cleaning denotes SHA-256-based duplicate image detection and removal prior to dataset splitting. (ii) The proposed framework uniquely combines dataset cleaning, attention-guided lesion localization, Smart Zoom ROI extraction, dual-stream feature learning, and Explainable AI within a unified pipeline.

7 Discussion

The results show that the proposed model gives better performance for potato leaf disease classification. The ablation study shows that every module helps to improve the accuracy. The attention module helps the model to focus on the infected part of the leaf, and the Smart Zoom layer selects the important region automatically. The cross-dataset results show that the model performs very well on the same dataset, but the accuracy drops when another dataset is used. This happens because the images are different in terms of lighting, background, and disease appearance. Therefore, testing on different datasets is important to check the practical performance of the model. The dataset was also checked for duplicate images using SHA-256 hashing. Duplicate images were removed, and the training, validation, and testing sets were verified to make sure that there was no overlap between them. This helped avoid data leakage. The proposed framework achieved better results than the existing methods on the PLD_Indonesia dataset. It combines attention maps, Smart Zoom module, dual-stream feature learning, and Explainable AI in a single model.

7.1 Limitations

The proposed architecture still has some limitations. The cross-dataset results show that domain shift remains a challenge when the model is tested on different datasets. The PLD_Indonesia dataset is also imbalanced, which may reduce the prediction performance for classes with fewer samples. The experiments were conducted only on publicly available datasets from a few geographical regions. Therefore, the model should be tested on more field images collected under different climatic conditions.

7.2 Conclusion

This study presents a Dual-Stream Center-Weighted Smart Zoom Architecture for potato leaf disease classification. The proposed architecture uses a Center-Weighted Multi-Scale Attention Module, a weakly supervised Smart Zoom layer, a shared EfficientNetV2B0 backbone, and a SHA-256-based dataset preparation pipeline. The model was evaluated on the PlantVillage, PLD_Pakistan, and PLD_Indonesia datasets and showed good performance on all three datasets. On the challenging PLD_Indonesia dataset, it achieved an accuracy of 88.12%, a balanced accuracy of 88.25%, and an MCC of 0.8555. The Explainable AI results showed that the model mainly focused on the disease-affected regions while making its predictions.

7.3 Future Work

In future work, self-supervised and semi-supervised learning methods will be explored to reduce the need for large labeled datasets. Vision Transformer and hybrid CNN-Transformer models will also be studied to improve feature learning. The Smart Zoom mechanism can be extended for disease segmentation and disease severity estimation. In addition, lightweight model compression and quantization techniques will be used to support deployment on mobile, edge, and IoT-based agricultural devices. The proposed architecture will also be tested on larger field datasets collected from different countries to evaluate its performance under real-world conditions.

References

- [1] Food and Agriculture Organization of the United Nations: A guide to the International Day of Potato. <https://openknowledge.fao.org/handle/20.500.14283/cd4980en>. Accessed: 2025-08-24 (2025)
- [2] Haverkort, A.J., Struik, P.C., Visser, R.G.F., Jacobsen, E.: Applied biotechnology to combat late blight in potato caused by *phytophthora infestans*. *Potato Research* **52**(3), 249–264 (2009) <https://doi.org/10.1007/s11540-009-9136-3>
- [3] Mozumder, A.I., Suman, R.M., Masood, R.U.: Comparative analysis of efficient face recognition models: A case study of efficientnetv2b0, nasnetmobile, densenet169, mobilenet v2 with transfer learning, l2 regularization and dropout. *SSRN Electronic Journal* (2025) <https://doi.org/10.2139/ssrn.5552618>
- [4] Fathonah, Z.L., Adi, P.W., Wibawa, H.A.: Optimized multiclass classification of skin lesions using a modified efficientnetv2b0 with transfer learning. *Discover Artificial Intelligence* (2026) <https://doi.org/10.1007/s44163-026-01545-0>
- [5] González-Briones, A., Florez, S.L., Chamoso, P., Castillo-Ossa, L.F., Corchado, E.S.: Enhancing plant disease detection: Incorporating advanced cnn architectures for better accuracy and interpretability. *International Journal of Computational Intelligence Systems* **18**(1), 120 (2025) <https://doi.org/10.1007/s44196-025-00835-2>

- [6] J, K., Natarajan, V.A.: A hybrid cbam and coordinate attention framework integrated with efficientnet for multi-class sugarcane leaf disease detection. SSRN Electronic Journal (2026) <https://doi.org/10.2139/ssrn.6000552>
- [7] Nurullah, M., Hodhod, R., Carroll, H., Zhou, Y.: Advancing multi-label tomato leaf disease identification using vision transformer and efficientnet with explainable ai techniques. Electronics **14**(23), 4762 (2025) <https://doi.org/10.3390/electronics14234762>
- [8] Alhammad, S.M., Khafaga, D.S., El-Hady, W.M., Samy, F.M., Hosny, K.M.: Deep learning and explainable ai for classification of potato leaf diseases. Frontiers in Artificial Intelligence **7**, 1449329 (2025) <https://doi.org/10.3389/frai.2024.1449329>
- [9] Hughes, D.P., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. CoRR **abs/1511.08060** (2015) [1511.08060](https://arxiv.org/abs/1511.08060)
- [10] Shabrina, N.H., Indarti, S., Maharani, R., Kristiyanti, D.A., Irmawati, Prastomo, N., Adilah M, T.: A novel dataset of potato leaf disease in uncontrolled environment. Data in Brief **52**, 109955 (2024) <https://doi.org/10.1016/j.dib.2023.109955>
- [11] Rashid, J., Khan, I., Ali, G., Almotiri, S.H., AlGhamdi, M.A., Masood, K.: Multi-level deep learning model for potato leaf disease recognition. Electronics **10**(17) (2021) <https://doi.org/10.3390/electronics10172064>
- [12] Nazir, T., Iqbal, M.M., Jabbar, S., Hussain, A., Albathan, M.: Efficientnet—an optimized and efficient deep learning approach for classifying disease of potato plant leaves. Agriculture **13**(4) (2023) <https://doi.org/10.3390/agriculture13040841>
- [13] Nasra, P., Singh, J., Rani, S., Shandilya, G., Bharany, S., Sood, S., Rehman, A.U., Hussien, S.: Optimized rexnet variants with spatial pyramid pooling, coordattention, and convolutional block attention module for money plant disease detection. Discover Sustainability **6**(1), 391 (2025) <https://doi.org/10.1007/s43621-025-01241-6>
- [14] Wahyunigrum, R.T., Siradjuddin, I.A., Farawati, T.D., Faradisa, I.S., Bauravindah, A., Dewi, D.A.: Implementation of an efficientnet-b4 model architecture with a convolutional block attention module (cbam) for betel leaf disease classification. Engineering, Technology & Applied Science Research **15**(5), 26722–26730 (2025) <https://doi.org/10.48084/etasr.11900>
- [15] Hoque, M.D.J., Islam, M.S.: An explainable and lightweight cnn framework for robust potato leaf disease classification using grad-cam visualization. Applied AI Letters **7**(1), 70011 (2026) <https://doi.org/10.1002/ail2.70011>

- [16] Sinamenye, J.H., Chatterjee, A., Shrestha, R.: Potato plant disease detection: leveraging hybrid deep learning models. *BMC Plant Biology* **25**(1), 647 (2025) <https://doi.org/10.1186/s12870-025-06679-4>
- [17] Charisma, R.A., Dharma Adhinata, F.: Transfer learning with densenet201 architecture model for potato leaf disease classification. In: 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE), pp. 738–743 (2023). <https://doi.org/10.1109/ICCoSITE57641.2023.10127772>
- [18] Saha, A., Musharraf, S.M., Dey, A., Roy, H., Bhattacharjee, D.: Potato leaf disease detection using cnn – a lightweight approach. In: Proceedings of the 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024). CEUR Workshop Proceedings, vol. 3900, pp. 158–171 (2024). Published under CC BY 4.0 license. <https://ceur-ws.org/Vol-3900/Paper13.pdf>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

