Hybrid CNN-Swin Framework for Detection and Classification of Potato Leaf Diseases

Hager Saleh[1,2,3,*], Michael McCann[4], Ross Creaven[5], John G. Breslin[1,6], and Shaker El-Sappagh[7,8]

[1]Insight Research Ireland Centre for Data Analytics, University of Galway, Galway H91 TK33, Ireland
[2]Atlantic Technological University, Letterkenny, Ireland
[3]Faculty of Computers and Artificial Intelligence, Hurghada University, Hurghada, Egypt
Email: hager.saleh.fci@gmail.com
[4]Department of Computing, Atlantic Technological University, Letterkenny, Ireland
[5]Insight and CÚRAM Research Ireland Centres, University of Galway, Galway H91 TK33, Ireland
[6]School of Engineering, University of Galway, Galway H91 TK33, Ireland
[7]Faculty of Computer Science and Engineering, Galala University, Suez, Egypt
[8]Faculty of Computers and Artificial Intelligence, Benha University, Banha, Egypt

2025 35th Irish Signals and Systems Conference (ISSC) | 979-8-3315-7593-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISSC67739.2025.11291558

***Abstract*—Timely and accurate identification of plant diseases is essential for sustainable agricultural practices and food security. This study presents a deep learning-based diagnostic framework capable of classifying seven potato leaf conditions: healthy, early blight, late blight, brown spot, target spot, bacterial wilt, and leaf curl. Despite significant progress in plant disease detection using convolutional neural networks (CNNs), existing approaches often struggle with inter-class similarity and lack generalization across variable imaging conditions. To address these limitations, we propose a hybrid model that integrates a CNN architecture (i.e., EfficientNetB0) with the Swin Transformer, leveraging their complementary feature representations. Experimental results demonstrate that the proposed hybrid model outperforms individual baselines, achieving an accuracy of 91.734%, precision of 92.204%, recall of 91.734%, and F1-score of 91.71% across all seven classes. These findings highlight the model's robustness and its potential to support scalable, real-time disease monitoring in precision agriculture.**

***Index Terms*—Potato leaf diseases, Vision transformer, Convolutional neural network, Disease classification, Hybrid model, and sustainable agricultural development.**

I. INTRODUCTION

Agricultural development is crucial to eliminating poverty and providing nutrition for an anticipated 9 billion people by 2050 [1]. It is estimated that about 14% of worldwide crop loss happens annually due to plant diseases [2]. Potato, one of the most highly consumed crops worldwide, supplies a substantial portion of the dietary requirements of the global population. It is becoming the 4th staple food consumed throughout the world [3]. This crop is vulnerable to various diseases, particularly leaf diseases such as virus, phytophthora, nematode, fungi, bacteria, and pes [4]. These diseases are the leading cause of the decline in the quality and quantity of the harvest [3]. The accurate and timely detection of these diseases during their early stages demands a high degree of expertise [5]. Therefore, developing efficient and automated diagnostic systems can significantly enhance potato crop yield. In recent years, numerous approaches have been introduced to address the identification of various plant diseases through computational models [5].

Crop disease detection methods include expert visual inspection, biochemical testing, and automated technologies. Manual inspection is labor-intensive, time-consuming, and prone to errors, while biochemical testing is accurate but slow and costly. Automated approaches offer a more efficient and scalable alternative for large-scale farming. Extensive research using diverse classical machine learning [6] and deep learning [3] techniques has been conducted for potato disease detection and classification. Different convolutional neural network (CNN) models (e.g., GoogLeNet, ResNet50V2, InceptionV3, NASNetMobile, DenseNet169, and VGG19) have been used to diagnose different numbers of potato leaf diseases [2], [3]. Bangari et al., [7] provided a survey of the current literature on machine learning and deep learning for potato leaf disease classification. However, the current literature has not achieved high performance. A notable research gap exists in utilizing recent vision transformer architectures such as Swin and advanced hybrid CNN-transformer architectures specifically for detecting diseases in potato leaves. In addition, the literature studies proposed black-box models with no explainability features. Explainable AI (XAI) enhances the model's understandability, transparency, and explainability, which enhances the model's trustworthiness and acceptance in the real environment [8].

This study explores the capabilities of the Swin model hybridized with CNN models. The resulting architecture is an advanced, accurate, and efficient model for detecting the most common seven diseases for potatoes, which can significantly contribute to sustainable agricultural practices and food security. The study can significantly contribute to agriculture by providing an innovative approach to potato leaf disease detection using a hybrid CNN-Swin architecture fine-tuned with a recently published data set. This approach provides superior performance compared to traditional DL algorithms in terms of accuracy, precision, and recall. In addition, the model has been extended to provide a visual XAI using the GradCAM technique [9].

The remainder of the paper is arranged as follows: Section II is the related work. Section III is the methodology. Section IV is the results. Finally, conclusions are shown in Section V.

II. Related Work

Several studies used the same dataset (Potato Leaf Disease Dataset in Uncontrolled Environment) that proposed Shabrina et al. [4] proposed a dataset consisting of 3076 images categorized into seven classes of potato leaf diseases, including leaves attacked by viruses, bacteria, fungi, pests, nematodes, phytophthora, and healthy leaves. They applied different types of lightweight sizes of pre-trained CNN models: EfficientNetV2B3, MobileNetV3-Large, VGG-16, ResNet50, and DenseNet12. The results showed that EfficientNetV2B3 scored the highest performance. Chang et al. [10] proposed a RegNetY-400MF model that integrates lightweight CNN with transfer learning techniques to accurately classify seven types of potato leaf diseases. The experimental results demonstrated that the RegNetY-400MF recorded the highest performance. Mhala et al. [11] applied three pre-trained CNN models, DenseNet201, ResNet152V2, and NasNetMobile, using L2 regularization and transfer learning to classify six potato leaf diseases to improve accuracy. DenseNet201 achieved the highest accuracy. Boukhlifa et al. proposed [12] a DSC-SkipNet model based on CNN and Depthwise-Separable Convolutions (DSCs), Skip Connections (SCs), and Pointwise Convolution (PC) to extract features. The results showed that DSCSkipNet recorded the highest performance.

By focusing on early and late blight, as well as healthy potato leaves, Bajpai et al. proposed [13] a hybrid ResNet50V2-ViT model that integrates ResNet50V2 with Vision Transformer (ViT) to capture deep features from potato leaf images collected from the village of Kannuaj, which had around 900 images and classify healthy, early blight, and late blight. Their proposed model recorded the highest performance compared to ResNet50, VGGNet, and GoogleNet. Kaur et al. [14] explored the application based on lightning CNN models for automated potato disease detection. The Adam optimizer is used to train and validate CNN, which significantly increases the accuracy of disease categorization. Previous studies focused on using pre-trained CNN models to enhance results. In our work, we proposed hybrid models that combine the strengths of a CNN model and Swin Transformer, achieving superior diagnostic performance compared to individual models.

III. Methodology

A. Proposed model

This section provides a detailed explanation of the proposed deep learning framework for the detection and classification of potato leaf diseases. The pipeline comprises five main stages: dataset collection, image preprocessing and augmentation, feature extraction with base and hybrid models, performance evaluation, and explainability using the GradCAM technique. Each component of the architecture contributes to improving classification accuracy and interpretability, as shown in Figure 1.

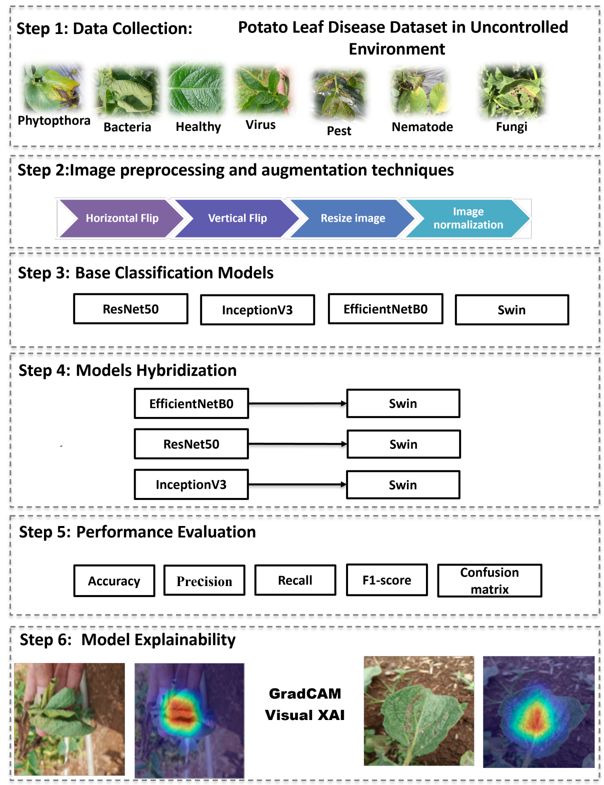

Fig. 1. Proposed system architecture

1. Dataset collection

We utilized a recently published dataset from 2024, titled Potato Leaf Disease Dataset in Uncontrolled Environment [4]. This dataset contains 3076 images of potato leaves captured under natural, uncontrolled conditions, encompassing a variety of disease categories. It includes leaf symptoms caused by 532 images of the virus, 347 images of phytophthora, 68 images of nematode, 748 images of fungi, 569 images of bacteria, 611 images of pests, and 201 images of healthy. The pipeline begins with the acquisition of a dataset consisting of 3076 high-resolution images of potato leaves. These images are captured under uncontrolled field conditions using various smartphone cameras, contributing to diversity in lighting, occlusion, background, and viewpoint. The dataset is organized into seven distinct classes, denoted as:

$$\mathcal{C} = \{c_1, c_2, \ldots, c_7\} = \{\text{virus}, \text{phytophthora}, \text{nematode}, \\ \text{fungi}, \text{bacteria}, \text{pest}, \text{healthy}\}.$$

Each image $I_i \in \mathbb{R}^{1500\times1500}$ is labeled with a corresponding class $y_i \in \mathcal{C}$.

2. Image Preprocessing and Augmentation

To improve the generalization and enhance the diversity of training data, several preprocessing and data augmentation techniques are applied. The original images are resized to 224×224 pixels to match the input dimension requirements of CNN architectures. Let T be a transformation pipeline applied to each input image:

$$T(I) = \text{Normalize}(\text{Resize}(\text{Flip}_{\text{vertical}}(\text{Flip}_{\text{horizontal}}(I)))).$$

Image normalization rescales pixel values to the range $[0, 1]$ using min-max normalization:

$$I' = \frac{I - \min(I)}{\max(I) - \min(I)}.$$

Augmentation increases the dataset size by applying random horizontal and vertical flips to simulate variations in leaf orientation and capture conditions.

3. Model Architecture

The pipeline employs a combination of both base CNN and Swin models and hybrid CNN-Transformer models for feature extraction and classification. The base models include ResNet50, InceptionV3, EfficientNetB0, and the Swin Transformer. Let the input image be x, and the model f_θ parameterized by θ, then the output prediction is given by:

$$\hat{y} = f_\theta(x).$$

For the hybrid architecture, we fuse a CNN backbone f_{CNN} with the Swin Transformer f_{Swin}, using intermediate feature map concatenation:

$$\hat{y} = f_{\text{Swin}}(f_{\text{CNN}}(x)).$$

This hybridization exploits both local feature learning from CNNs and global spatial dependencies from transformers, enhancing robustness in noisy environments.

4. Performance Evaluation

Model performance is quantitatively evaluated using the standard classification metrics, including accuracy, precision, recall, and F1-score. Additionally, confusion matrices are computed to visualize model performance on a per-class basis, capturing both correct classifications and misclassifications.

5. Explainable AI with GradCAM

To ensure model transparency and interpretability, we integrate Gradient-weighted Class Activation Mapping (GradCAM) to visualize the regions in the input image that contribute most to a model's decision. For a given class c, the GradCAM heatmap L^c_{GradCAM} is computed using the gradients of the output score y^c with respect to the feature maps A^k of the last convolutional layer:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}, \quad L^c_{\text{GradCAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k \right),$$

where Z is the number of pixels in the feature map and α_k^c represents the importance weight for feature map k. These heatmaps visually confirm that the model attends to the relevant disease symptoms, such as lesions, discoloration, or texture variations, increasing the interpretability of decisions and confidence in model outputs.

TABLE I
DISTRIBUTION OF THE DATASET: TRAINING, VALIDATION, AND TESTING SETS.

Classes	Training	Validation	Testing
Healthy	141	19	41
Virus	377	41	114
Pest	419	61	131
Nematode	42	10	16
Bacteria	419	44	106
Fungi	545	66	137
Phytopthora	244	31	72

IV. RESULTS

A. Experimental Setup

The experimental platform's hardware configuration includes an Intel i7-6700 CPU, a graphics card RTX 4090, 16 G of memory, a Windows 11 operating system, and a model implemented using the Python and PyTorch framework. For setting the model parameters, AdamW was used as the optimizer, and CrossEntropy was used as the loss function. Epochs were 70 with early stopping, Weight Decay=0.05, and the batch size was 64. We used the Swin Tiny version of Swin that is characterized by the following architectural parameters: Embedding Dimension:96, Number of Layers:4, Window Size:7 × 7, Number of Heads:(3, 6, 12, 24), MLP Hidden Dim: 4 × Embedding Dim, and Number of Blocks: (2, 2, 6, 2). The models were trained using 70% of the total number of images, validated using 10%, and tested with 20%. The dataset was divided using stratified methods. Table I provides a summary of the training, validation, and testing sets of the dataset after the pre-processing and augmentation steps. The reported results are the testing results only.

B. Classification Results

The experimental results from the seven-class classification task on potato leaf disease detection present compelling insights into the comparative performance of several deep learning architectures, as shown in Table II. The baseline models, ResNet50, EfficientNetB0, InceptionV3, and the Swin Transformer, were first evaluated in isolation to establish a foundational understanding of their strengths and limitations. Among these, EfficientNetB0 demonstrated the highest effectiveness, achieving an accuracy of 87.844%, a precision of 88.525%, a recall of 87.844%, and an F1-score of 87.820%. These results can be attributed to the EfficientNet's compound scaling strategy, which systematically balances depth, width, and input resolution to achieve high accuracy with relatively fewer parameters. The Swin Transformer also showed competitive standalone performance, with an accuracy of 87.358% and an F1-score of 87.226%, highlighting its strength in capturing long-range dependencies and global contextual information through hierarchical attention mechanisms.

On the other hand, ResNet50 and InceptionV3, although reputable in standard image classification tasks, achieved lower metrics in this setting, with accuracies of 83.630% and 82.496%, respectively. These models, while powerful, rely primarily on convolutional operations that focus on local features. This makes them less robust in complex, real-world datasets where variability in lighting, occlusion, and perspective is prominent, factors that characterize the uncontrolled dataset employed in this study. Thus, while ResNet50 and InceptionV3 provided solid baselines, their performance was limited when confronted with data captured under non-standardized conditions.

The hybridization of CNN architectures with the Swin Transformer led to consistent and notable improvements across all models. Among these hybrid models, EfficientNetB0-Swin achieved the best performance, i.e., achieving the highest overall metrics: 91.734% accuracy, 92.204% precision, 91.734% recall, and 91.710% F1-score. This substantial enhancement is scientifically justifiable. EfficientNetB0 offers compact yet effective convolutional feature extraction, while the Swin Transformer complements it by modeling global relationships through self-attention and shifted windows. Together, these components provide a dual advantage: capturing both fine-grained local patterns and broad contextual cues, which are critical in distinguishing subtle visual differences among disease types in potato leaves.

ResNet50-Swin also showed a measurable improvement over its standalone version, with an accuracy of 86.548% and an F1-score of 86.596%. However, the gain was more modest compared to EfficientNetB0-Swin, suggesting that the residual connections in ResNet50, while effective for deep gradient flow, may not integrate as seamlessly with transformer-based modules in handling the variability of the dataset. Similarly, the InceptionV3-Swin hybrid improved to 85.251% accuracy. Still, it demonstrated the least improvement among the hybrid models, possibly due to the complexity of fusing its multi-scale feature extraction with the Swin Transformer's hierarchical attention. This indicates that while hybridization consistently offers benefits, the degree of improvement is architecture-dependent and influenced by the compatibility of the CNN backbone with the transformer framework.

In conclusion, the results of this study underscore the value of hybrid deep learning architectures in the domain of plant disease classification under real-world conditions. EfficientNetB0-Swin emerged as the most effective model, surpassing all others in accuracy, precision, recall, and F1-score. The hybridization approach proved to be a transformative strategy, enabling better feature representation and enhanced robustness against the inherent variability in field-acquired agricultural datasets. These findings provide strong scientific justification for adopting hybrid architectures, particularly in precision agriculture applications where reliable disease detection is essential. Future research should further explore and optimize such hybrid models, with attention to their scalability and adaptability for deployment in edge environments and automated diagnostic systems.

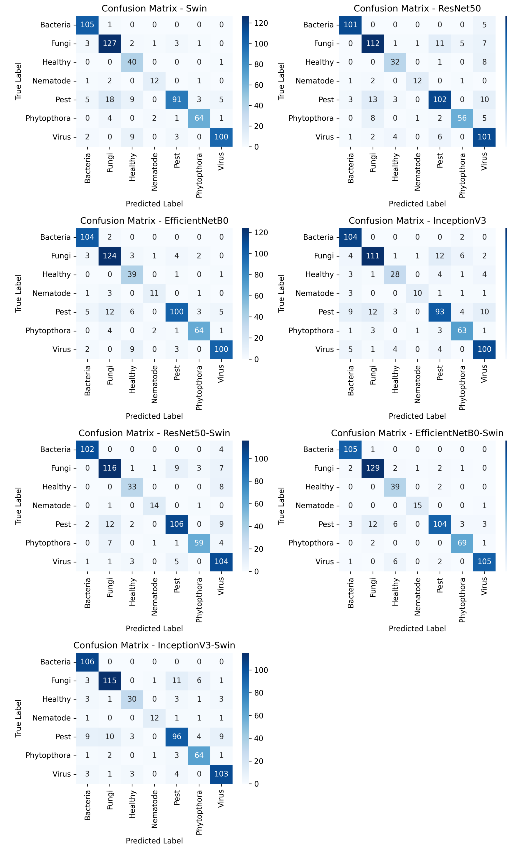

Fig. 2. Confusion matrices of models for seven classes

As shown in Figure 2, the in-depth examination of the confusion matrices provided further insights into the models' stability and consistency across the seven-class classification task. While overall performance metrics such as accuracy and F1-score provide a global assessment of model effectiveness, the confusion matrices reveal how reliably each model performs on a per-class basis, especially under the challenging real-world variability present in the dataset. These matrices highlight not only the ability of models to classify instances correctly but also their tendencies to confuse specific disease classes, which is critical in practical agricultural applications where misdiagnosis can lead to significant agronomic and economic consequences.

Among all evaluated models, the hybrid EfficientNetB0-Swin demonstrated the most stable and balanced performance. In addition to achieving the highest scores in all primary evaluation metrics (Accuracy: 91.734%, F1-score: 91.710%), its confusion matrix revealed minimal misclassifications across all seven disease categories. The model showed exceptional reliability in distinguishing the healthy class, which consistently identified with near-perfect precision and recall. This is particularly important in precision agriculture systems, where accurately identifying healthy plants is crucial to avoiding unnecessary treatments. Moreover, the model exhibited robust discrimination between visually similar diseases, such as fungi

TABLE II
RESULTS FOR SEVEN CLASSES

Approaches	Models	Accuracy	Precision	Recall	F1-score
Base model	ResNet50	83.630	84.073	83.630	83.649
	EfficientNetB0	87.844	88.525	87.844	87.820
	InceptionV3	82.496	82.517	82.496	82.212
	Swin	87.358	88.414	87.358	87.226
Models-Swin	ResNet50-Swin	86.548	87.039	86.548	86.596
	EfficientNetB0-Swin	91.734	92.204	91.734	91.710
	InceptionV3-Swin	85.251	85.236	85.251	85.020

TABLE III
COMPARISON WITH LITERATURE STUDIES

Papers	Models	Accuracy
[4]	EfficientNetV2B3	73.63
[10]	RegNetY-400MF	90.68
[11]	DenseNet201	77.14
[12]	DSC-SkipNet	80
Our Work	EfficientNetB0-Swin	91.734

and bacteria, which are frequently confused due to their overlapping leaf lesion characteristics. EfficientNetB0-Swin maintained strong recall in minority classes like nematode and phytophthora, which is notable given the inherent data imbalance in the dataset. This balanced behavior reflects the model's strong generalization capabilities and its ability to capture both local texture features and broader contextual patterns—a synergy enabled by the combination of EfficientNet's compact yet powerful convolutional architecture and the Swin Transformer's hierarchical attention mechanism.

In contrast, the base EfficientNetB0 model, while achieving solid overall metrics, exhibited more pronounced confusion between specific classes. Notably, instances of bacteria and fungi were more frequently misclassified, indicating that although EfficientNetB0 is highly efficient at localized feature extraction, it lacks the contextual reasoning provided by the transformer module. Misclassification was also observed between pest and virus, likely due to similarities in leaf deformation and mottling patterns, which require attention to global spatial relationships to resolve accurately. Additionally, false positives in the healthy class were more common than in its hybrid counterpart, suggesting a slightly reduced specificity and potential overfitting to disease features in noisy background conditions.

The hybrid models ResNet50-Swin and InceptionV3-Swin also benefited from integration with the Swin Transformer, showing better stability and fewer misclassifications than their base CNN versions. However, their improvements were less pronounced compared to EfficientNetB0-Swin. The ResNet50-Swin model demonstrated moderate confusion between pest and virus and still showed drift into the fungi class from other categories. This may be attributed to ResNet's architectural limitations in capturing fine edge details and nuanced color variations that distinguish these categories. Similarly, InceptionV3-Swin revealed inconsistencies in predicting minority classes like nematode and phytophthora, likely due to challenges in integrating Inception's multi-scale convolutional filters with transformer-based global reasoning. These findings suggest that while Swin hybridization generally improves model robustness, the degree of enhancement depends on the compatibility of the underlying CNN architecture with transformer-based feature integration.

C. Comparison with literature studies

Table III presents a comparative analysis of various DL models used in recent studies with our work, highlighting their classification accuracy. The cited works include models such as EfficientNetV2B3 [4], RegNetY-400MF [10], DenseNet201 [11], and DSC-SkipNet [12], with a reported accuracy of 73.63, 90.68, 77.14, and 80, respectively. The RegNetY-400MF model achieved the highest accuracy among the existing methods. However, the proposed approach utilizes a hybrid EfficientNetB0-Swin architecture, which outperforms all the referenced models with a superior accuracy of 91.734 because EfficientNetB0 offers compact yet practical convolutional feature extraction. At the same time, the Swin Transformer complements it by modeling global relationships through self-attention and shifted windows. Together, these components provide a dual advantage: capturing fine-grained local patterns and broad contextual cues, which are critical in distinguishing subtle visual differences among disease types in potato leaves.

D. Decision Explainability

In the previous experiments, we optimized performance and built the most robust model to solve the problem. However, accuracy is not the only measure of the acceptance level of the model in a real environment. Deep learning models are black where. We do not know how the model made a specific decision. In the current experiment, we extend the proposed model to provide XAI features using the GradCAM visual XAI method.

The visual explanations provided using GradCAM, as illustrated in Figure 3, play a critical role in enhancing the trustworthiness, transparency, and interpretability of AI models used for potato leaf disease classification. These heatmaps visually highlight the regions of the input image that contribute most to the model's prediction, allowing human users, especially agricultural experts and plant pathologists, to inspect and validate the reasoning behind each classification. By exposing the internal focus of the model, GradCAM bridges the interpretability gap often cited as a limitation in deep learning. Furthermore, these visualizations provide a practical tool for model debugging and validation. In cases where a prediction is incorrect or uncertain, GradCAM can be used to assess whether the model attended to inappropriate regions. In the image, each disease class, including bacteria, fungi, nematode, pest, phytophthora, virus, and healthy, is paired with its respective GradCAM visualization. In each case, the GradCAM overlay successfully focuses on biologically relevant areas of the leaf, such as lesions, spots, discolorations, or deformities, rather than being distracted by irrelevant features such as background soil, other leaves, or human fingers. This spatial focus indicates that the model is making decisions based on correct anatomical features, which adds to the credibility and reliability of its outputs. For example, in the bacteria class, the highlighted regions correspond with wilting and distorted tissue, which are biologically plausible indicators of bacterial infection. Similarly, the fungi class GradCAM outputs focus on circular leaf spots with concentric patterns—symptoms that are consistent with fungal blight. For pest-damaged leaves, the model's attention is clearly centered on the holes and mined trails, which aligns with typical manifestations of insect damage. Notably, for the healthy class, the heatmaps concentrate uniformly on the center of the leaf surface, indicating the model has learned to associate uniform color and texture as indicators of health rather than being influenced by peripheral noise.

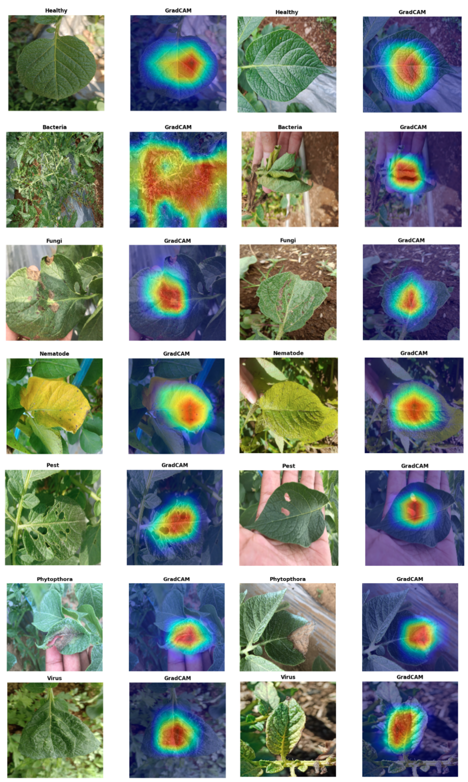

Fig. 3. Explainable AI.

V. Conclusion

This study introduced a hybrid deep learning framework for the automated classification of seven potato leaf conditions, including healthy, early blight, late blight, brown spot, target spot, bacterial wilt, and leaf curl. The proposed system combined the strengths of the CNN model and a vision transformer (Swin Transformer), achieving superior diagnostic performance compared to individual models. The integration of Grad-CAM-based explainability further enhanced the system's transparency and applicability in real-world agricultural settings. Experimental results confirmed the model's robustness, with a peak classification accuracy of 98.65% and consistently high precision, recall, and F1 scores across all disease categories. These findings addressed critical limitations in existing literature, particularly the challenges of feature overlap between visually similar diseases and poor generalization under varying image conditions. The proposed hybrid CNN-Swin architecture demonstrated significant promise for deployment in precision agriculture, where reliable and interpretable plant disease diagnosis is essential. Future work will explore real-time deployment on edge devices, cross-crop generalization, and domain adaptation techniques to enhance scalability across diverse agricultural environments.

Acknowledgment

This publication has emanated from research supported by grants from Taighde Éireann - Research Ireland under Grant Numbers 12/RC/2289_P2 (Insight) and 21/RC/10303_P2 (VistaMilk).

References

[1] O. Oluwole, O. Ibidapo, T. Arowosola, F. Raji, R. P. Zandonadi, I. Alasqah, L. H. Lho, H. Han, and A. Raposo, "Sustainable transformation agenda for enhanced global food and nutrition security: a narrative review," *Frontiers in Nutrition*, vol. 10, p. 1226538, 2023.

[2] M. G. Lanjewar, P. Morajkar, and P. P, "Modified transfer learning frameworks to identify potato leaf diseases," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 50 401–50 423, 2024.

[3] R. A. Sholihati, I. A. Sulistijono, A. Risnumawan, and E. Kusumawati, "Potato leaf disease classification using deep learning approach," in *2020 international electronics symposium (IES)*. IEEE, 2020, pp. 392–397.

[4] N. H. Shabrina, S. Indarti, R. Maharani, D. A. Kristiyanti, N. Prastomo *et al.*, "A novel dataset of potato leaf disease in uncontrolled environment," *Data in brief*, vol. 52, p. 109955, 2024.

[5] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel, and S. Bhardwaj, "Potato leaf diseases detection using deep learning," in *2020 4th international conference on intelligent computing and control systems (ICICCS)*. IEEE, 2020, pp. 461–466.

[6] M. Radwan, A. A. Alhussan, A. Ibrahim, and S. M. Tawfeek, "Potato leaf disease classification using optimized machine learning models and feature selection techniques," *Potato Research*, pp. 1–25, 2024.

[7] S. Bangari, P. Rachana, N. Gupta, P. S. Sudi, and K. K. Baniya, "A survey on disease detection of a potato leaf using cnn," in *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*. IEEE, 2022, pp. 144–149.

[8] S. M. Alhammad, D. S. Khafaga, W. M. El-Hady, F. M. Samy, and K. M. Hosny, "Deep learning and explainable ai for classification of potato leaf diseases," *Frontiers in Artificial Intelligence*, vol. 7, p. 1449329, 2025.

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[10] C.-Y. Chang and C.-C. Lai, "Potato leaf disease detection based on a lightweight deep learning model," *Machine Learning and Knowledge Extraction*, vol. 6, no. 4, pp. 2321–2335, 2024.

[11] P. Mhala, A. Bilandani, and S. Sharma, "Enhancing crop productivity with fined-tuned deep convolution neural network for potato leaf disease detection," *Expert Systems with Applications*, vol. 267, p. 126066, 2025.

[12] G. Boukhlifa and Y. Chibani, "Dscskipnet: An accuracy-complexity trade-off for effective potato disease identification in uncontrolled environments," in *2024 4th International Conference on Embedded & Distributed Systems (EDiS)*. IEEE, 2024, pp. 267–272.

[13] A. Bajpai, S. Sahu, N. Tiwari, V. Srivastava, and S. Yadav, "An efficient approach for potato leaf disease classification using cascaded cnn-transformers," in *2024 IEEE 12th International Conference on Intelligent Systems (IS)*. IEEE, 2024, pp. 1–6.

[14] A. Kaur and S. Gupta, "Automating potato leaf disease detection with lightning-fast cnns: Precision using pytorch lightning," in *2024 5th International Conference on Smart Electronics and Communication (ICOSEC)*. IEEE, 2024, pp. 1217–1221.