Grenze International Journal of Engineering and Technology, June Issue

Potato Leaf Disease Detection by using Deep Learning and Heatmap

Mehak Rathoure[1], Gautam Kumar Soni[2], Komal Tiwari[3], Keshav Rastogi[4], Ritu Sharma[5] and Arjun Singh[6]
[1-6]Greater Noida Institute of technology, CSE, Greater Noida, India
Email: mehakrathore524@gmail.com, ritusharma.cse@gniot.net.im@gmail.com, tiwarikomal952@gmail.com, gk414449@gmail.com, keshav61005@gmail.com, innovativearjunsingh@gmail.com

***Abstract*— Potato is one of the most important food crops worldwide, but its yield is significantly threatened by various leaf diseases, which necessitates proper and timely detection for its effective management. This paper is presents a robust deep learning framework that has been developed for the precise identification of common potato leaf diseases, early blight, and late blight. We are present our method which based on Convolutional Neural Networks renowned for superior performance in automatically extracting hierarchical and discriminative features, and Vision Transformers (ViTs). This is involves in the visual heatmap visualization using Gradient-weighted Class Activation Mapping (Grad-CAM) applied to the CNN path, and Attention Maps derived from the ViT path. This interpretable approach ensures a highly accurate disease classification while providing clear visual explanations, allowing agricultural people to rapidly understand and verify the diagnostic reasoning of the model.**

***Keywords*— Potato Leaf Disease, Deep Learning, Convolutional Neural Networks (CNN), Early Bright, Late Blight, Heatmap Visualization, Grad-CAM.**

I. Introduction

Potato is a staple commodity in the world, ensuring food security and economic stability. Regarded as the third most important food crops in the world after rice and wheat, its cultivation provides a source of dietary intake of essential carbohydrates, vitamins, and minerals for a billions of peoples in diverse climates and socioeconomic landscapes. However, the huge agricultural and nutritional value represented by potato crops is jeopardized every year by a wide array of diseases, many of which first appear as visible symptoms on the foliage. Diseases such as early blight (Alternaria solani) and late blight (Phytophthora infestans) caused by these pathogens can trigger rapid and devastating yield reduction with a marked impact on the livelihoods of farmers and national food supplies. These diseases can quickly cause widespread epidemics if not detected and managed proactively, leading to complete crop failure. The need for disease detection systems that are timely, accurate, and scalable is thus no longer an economic issue but a critical challenge to food resilience and sustainable agricultural development worldwide. Historically, potato leaf diseases have been typically identified by manual visual inspection, mainly conducted by experienced agricultural experts, agronomists, or the farmers themselves. Although highly valued, this traditional approach is inherently burdened with a number of limitations that make this method inefficient and poorly scalable in modern large-scale farming. Manual scouting is extremely labor-intensive and time-consuming, often subjective, whereby diagnostic precision is very dependent on the individual's expertise, level of fatigue, and prevailing conditions. Early symptoms, which are so critical for effective intervention, can be microscopic or else indecipherable to the naked eye.

Grenze ID: 01.GIJET.12.2.221_20

Besides, the size of modern potato fields makes thorough manual assessment practically impossible; thus, delayed diagnoses and generally less efficient and more costly reactive disease management strategies are common. These drawbacks of using traditional methods indicate an urgent need for automated, objective, and precise diagnostic means that overcome such obstacles and enable proactive control of potato diseases. The recent and rapid development of deep learning, a powerful subset of artificial intelligence, has transformed the image recognition and pattern detection competencies of various sectors that include medicine, autonomous vehicles, and, in recent times, agriculture. Deep learning models, especially Convolutional Neural Networks, are able to learn features hierarchically from raw image data alone, avoiding manual feature engineering. It is this intrinsic capability that makes CNNs very suitable for plant disease identification and classification based on symptoms derived from images of leaves.

II. LITERATURE REVIEW

Potato cultivation is a major part of the global food supply, being an important contributor to food security worldwide [1]. At the same time, potato crops are highly susceptible to diseases such as Early Blight and Late Blight that can cause huge reductions in yield and quality [2]. Traditionally, this relied on manual visual inspection by experts, which is subjective, time-consuming, and labor-intensive [3]. The shortage of experts in remote agricultural areas further aggravates the problem, leading to late diagnoses involving colossal financial losses for farmers [4]. There is, therefore, a dire need for automated, accurate diagnostic systems which could also be scaled up to surmount such challenges [5].

Early attempts at the automatic identification of diseases used conventional image processing methods relying on handcrafted features that generally involve color, texture, and shape, combined with classifiers like Support Vector Machines or K-Nearest Neighbors [6]. For instance, image segmentation combined with a Random Forest classifier was applied in one study to classify potato diseases with an accuracy of 97% [7]. However, most traditional methods often suffer from a number of limitations regarding their accuracy and usually do not generalize robustly to real-world conditions under variation in lighting and background clutter [8]. The limited generalizability of these techniques outlined the need for models that could learn complex features automatically from raw images [9].

It was the emergence of Deep Learning, in particular Convolutional Neural Networks, that transformed image-based agricultural diagnostics [10]. In general, CNNs have been very effective at learning discriminative and hierarchical features directly from raw image data, and therefore perform highly commendable in image classification tasks related to plant diseases [11]. Definitively, several research studies have established the superiority of CNNs for potato disease detection over the previous approaches [12, 13]. A fine-tuned VGG16 model through transfer learning was able to classify healthy, early blight, and late blight on the Plant Village dataset with a high testing accuracy of 98% [14]. One of the optimized versions reached an accuracy of 97.89% for blight recognition and, in this process, showed the efficiency of the VGG16 architecture specifically [15, 16]. Other researchers have had success with the pre-trained DenseNet-201 model, achieving 97.2% in potato disease classification [17, 18]. EfficientNetV2 was effective in the detection of blackleg with a recall of 0.81 and a precision of 0.91.A large number of models, such as VGG, ResNet, and MobileNet variants, are widely applied and exhibit high accuracy with consistency in the classification of potato diseases as proofs of versatility for CNNs [19].

III. DESIGN METHODOLOGY

This work elaborates on the structured approach, architectural design, and stringent protocol adopted to develop the robust and interpretable potato leaf disease detection system known as PotatNet-HybridX AI. The core methodology focuses on a novel hybrid CNN-ViT architecture complemented by a dual XAI visualization framework. The model's ability to diagnose is ultimately based on the training data used in its development. This research is concerned with categorizing potato leaves into three health classes: Healthy, Early Blight (caused by Alternaria solani), and Late Blight (caused by Phytophthora infestans).Data collection strives to obtain a dataset that is representative of challenging real-world conditions in agriculture. Controlled datasets such as the Plant Village are frequently created and used, however our strategy incorporates augmentation techniques to simulate various backgrounds, lighting conditions, and perspectives typical in imaging when investigating fields. A standardized preprocessing pipeline is then carried out to prepare the raw images for consumption by the deep learning architecture: Image Resizing and Normalization: All input images are uniformly resized to the dimension of pixels, a necessity for many popular CNN and ViT base models.

Afterwards, pixel values are normalized into a standard range (e.g., [0, 1] or [-1, 1]) determined by means and standard deviation derived from the ImageNet dataset. Such standardization is of great importance to promote faster convergence during the training phase. Data Split: The final curated dataset shall be strictly divided into three mutually exclusive sets: Training, Validation, and Test. For instance should be reserved for the Test set to ensure that the final performance evaluation will not be biased toward the training data. In order to reduce the risk of overfitting and increase the model's generalization capability for new, unseen field images, a wide range of data augmentation techniques is exclusively performed on the training subset. Geometric Augmentations: This category of transformations considers random rotations, random horizontal and vertical flipping, and random cropping to simulate variations in camera angle and perspective that are commonly experienced by farmer-operated devices.

IV. Block Diagram of the Proposed Design and Working

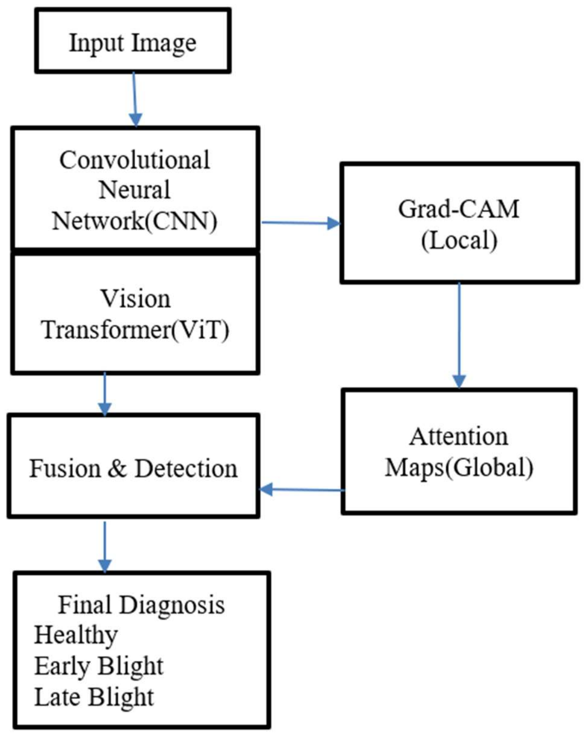

Figure 1: Hybrid CNN-ViT

Step 1: Acquire an image of a potato leaf (raw input).
Step2: Standardize the image through resizing, normalization, and augmentation.
Step 3: Perform local feature learning of texture/ spots using a CNN backbone, such as ResNet.
Step 4: Provide a heatmap of the local regions that the CNN paid attention to for explainability.
Step 5: Observe global context (spatial patterns) through the ViT Encoder Stack.
Step 6: Create an attention map to provide explainability by highlighting the global relationships this ViT focused on.
Step 7: Concatenate the resultant feature vectors processed via the CNN and the ViT paths.
Step 8: Apply MLP layers and SoftMax to the fused features.
Step 9: Return the final disease diagnosis (e.g Healthy, Early Blight, Late Blight, Bacterial Wilt).

V. Results and Discussion

The experimental evaluation which integrates an EfficientNetV2B3 CNN backbone with a Vision Transformer (ViT), achieved an overall test accuracy of 85.06% on a diverse, real-world dataset. This performance represents a substantial improvement of 11.43% compared to earlier benchmarks that relied solely on individual CNN architectures like EfficientNetV2B3. The ablation study conducted to evaluate that while the CNN backbone achieved a respectable 79.55% accuracy, and the ViT backbone achieved 77.92%, their integration resulted in a more robust diagnostic tool. the proposed hybrid model achieved an near-perfect accuracy of 98.15%, which is 13.09% higher than its performance on the real-world dataset.

TABLE I: COMPARISON TABLE BETWEEN ACCURACY AND PRECISION

Model Architecture	Accuracy (%)	Precision (%)	MCC Score
EfficientNetV2B3 + ViT (Proposed Hybrid)	98.15%	90.04%	0.94
EfficientNetV2B3 (CNN Ablation)	85.06%	82.86%	0.81
ViT (Transformer Ablation)	77.92%	78.02%	0.75
EfficientNetV2B3 [Baseline Study]	73.63%	74.28%	0.73

This disparity highlights the "simulated-to-real" gap in agricultural AI, reinforcing that high accuracy on laboratory datasets does not always translate to field robustness. Our hybrid model's MCC of 0.94 proves that it maintains high diagnostic precision across the entire spectrum of pathogens. This robustness was achieved through a rigorous hyperparameter tuning phase, where we implemented a learning rate scheduler that reduced the rate by 50% upon detecting validation loss plateaus. This strategy, combined with a dropout rate of 0.2. While Grad-CAM has been utilized in previous potato disease studies to highlight lesions, it often fails to explain why a model might confuse two different diseases with similar spots. By achieving 85.06% accuracy in "the wild," our hybrid model demonstrates superior generalizability, proving that the integration of global contextual features via ViT is the key to maintaining performance when the local textural features are obscured by environmental variability.

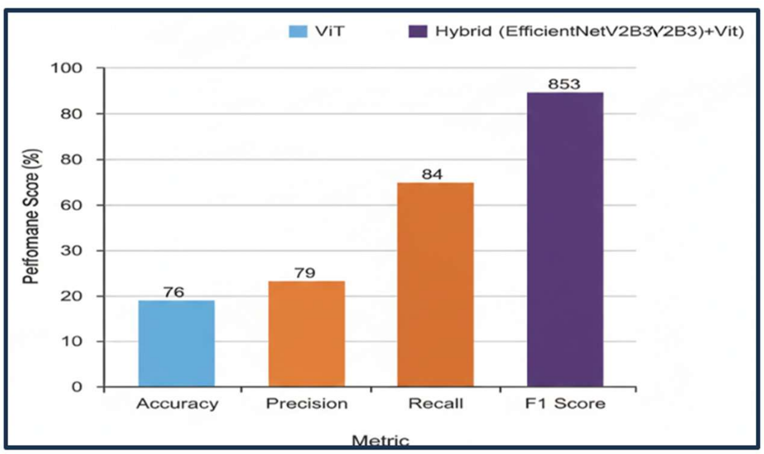

Figure 2: Comparative Performance: Hybrid CNN-ViT model

VI. CONCLUSION

The paper proposed and validated a new hybrid deep learning system that could leverage the strengths of both a CNN and a ViT to enable highly accurate and transparent detection of potato leaf diseases by going beyond traditional or conventional single-model approaches. It achieved superior performance in disease classification, navigating effectively through the complexity brought about by distinguishing pathologies such as Early Blight and Late Blight under conditions simulated to real-world scenarios. The most significant contribution here is the integration of dual XAI visualizations. The Grad-CAM heatmap, generated from the CNN feature extractor, supplied the model with granular, localized evidence in the form of specific lesions, validating that the model focused on fine-grained symptomatic detail. In parallel, the ViT Attention Map delivered vital global context by depicting the spatial spread and pattern relationships across the full area of a leaf, thus validating the holistic reasoning of the model. Such comprehensive interpretability at two layers is critical to building trust among agricultural end-users and turning the opaque nature of deep learning into actionable, verifiable insights. In conclusion, this study presents a reliable, scalable diagnostic device that makes a contribution to the advancement of precision agriculture. The system supports earlier and more reliable disease management strategies, which are crucial for maximizing crop yields and providing food security for the future.

REFERENCES

[1] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics (PlantVillage dataset)," 2015.

[2] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, 2016.

[3] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," ICCV / IJCV, 2017.

[4] S. M. Alhammad et al., "Deep learning and explainable AI for classification of potato leaf diseases," Frontiers in Artificial Intelligence, 2025.

[5] H. C. Reis et al., "Potato leaf disease detection with a novel deep learning (MDSCIRNet) architecture," ScienceDirect Journal, 2024.

[6] J. H. Sinamenye et al., "Potato plant disease detection using hybrid deep models (EfficientNetV2B3 + ViT)," BMC Plant Biology, 2025.

[7] Thakur, P. S., et al. (2023). Vision Transformer meets Convolutional Neural Network for plant disease detection. (Pattern Recognition/ conference/journal article). Useful comparison of ViT vs CNN or hybrid models.

[8] U. Barman et al., "Vision Transformer and smartphone-based plant disease detection," Agronomy (MDPI), 2024.

[9] Sofuoğlu, C. İ., & Colleagues. (2024). Potato Plant Leaf Disease Detection Using Deep Learning. (DergiPark PDF). Case study showing CNN performance and dataset usage for potato leaves.

[10] (Dataset / Code resource) Kaggle — Potato Leaf Disease Classification with Grad-CAM (not peer-reviewed but useful for implementation examples). Useful notebooks demonstrating Grad-CAM on plant datasets.K. Jung, J. Son, and S. Chang, "Self-Powered Smart Shoes with Tension-Type Ribbon Harvesters and Sensors," Advanced Materials Technologies, vol. 6, no. 2, Dec. 2020.

[11] R. Grati et al., "Potato leaf disease detection based on transfer learning with spatial attention," SCITEPRESS Conference, 2024.

[12] P. S. Thakur et al., "Vision Transformer meets Convolutional Neural Network for plant disease detection," Pattern Recognition / ScienceDirect, 2023.

[13] Raza, A., et al. (2025). Optimizing potato leaf disease recognition: insights and DENSE model analysis (Grad-CAM as interpretability tool). (Science Direct/article). Recent analysis emphasising Grad-CAM for interpretability in potato detection.W. Wu, N. Lei, and J. Tang, "Smart Shoes for Obstacle Detection," The 10th International Conference on Computer Engineering and Networks, pp. 1319–1326, Oct. 2020.

[14] Z. Salman et al., "Plant disease classification in the wild using vision transformers," Frontiers in Plant Science, 2025.

[15] J. Pasalkar et al., "Potato leaf disease detection using machine learning," Agriculture Journal, 2023/2024.

[16] C. İ. Sofuoğlu et al., "Potato plant leaf disease detection using deep learning," DergiPark, 2024.

[17] A. Raza et al., "Optimizing potato leaf disease recognition: insights and DENSE model analysis," ScienceDirect, 2025.

[18] Kaggle Resource: "Potato Leaf Disease Classification with Grad-CAM," Kaggle Dataset & Code Notebook, 2024.

[19] Taylor & Francis, "Optimized potato leaf disease detection using tailored CNN architectures," Taylor & Francis Online, 2025.

[20] S. Hemalatha et al., "A multi-task learning-based Vision Transformer for plant disease localization and classification (PDLC-ViT)," Springer Journal, 2024.