Lesion-Aware Explainability Validation for Apple Disease CNN using Grad-CAM and IoU

Sathwik parimi
IIT (ISM) Dhanbad,
Jharkhand, India
24je0956@iitism.ac.in

***Abstract*—Explainable AI (XAI) is critical in high-stakes domains such as agriculture, where understanding model decisions enhances reliability and adoption. This paper presents a lesion-aware explainability validation pipeline for CNN-based apple leaf disease classification. We utilize Grad-CAM to visualize the classifier's focus and validate it against lesion segmentation masks using Intersection-over-Union (IoU). Results show varying degrees of alignment, highlighting scenarios where the model's attention corresponds—or fails to correspond—to actual disease regions.**

***Index Terms*—Apple disease, CNN, Explainable AI, Grad-CAM, Intersection over Union, Lesion segmentation**

I. Introduction

The accurate and interpretable diagnosis of plant diseases is vital in agriculture. While deep convolutional neural networks (CNNs) have achieved high accuracy in image-based disease classification, their black-box nature remains a concern. This study introduces a framework to evaluate whether CNNs truly focus on pathological lesions during classification using Grad-CAM and lesion segmentation masks.

.

II. Realted work

- CNN-based Plant Disease Detection: CNNs have demonstrated high accuracy in plant disease classification tasks, but their interpretability remains limited.
- Attention Mechanisms: Recent models incorporate attention modules to enhance feature extraction from lesion areas, improving both accuracy and explainability.
- Explainability in Agriculture: Grad-CAM and similar XAI techniques are increasingly used to interpret model decisions. However, most studies rely on qualitative assessments, lacking rigorous quantitative validation against ground-truth lesion annotations.
- Quantitative Validation: Some recent works propose comparing Grad-CAM maps with lesion masks using IoU or Dice coefficients, providing a more objective measure of explainability.

III. Methodology

To validate whether a convolutional neural network (CNN) genuinely focuses on diseased lesions while classifying apple leaf diseases, we designed a pipeline combining image classification, explainability, and segmentation comparison. Each RGB image is paired with a binary ground-truth lesion mask. The images are preprocessed (resized, normalized) and classified using a ResNet-50 model, either pre-trained on ImageNet or fine-tuned on an apple disease dataset. Grad-CAM is used to generate class-specific heatmaps that visualize the spatial attention of the CNN. These heatmaps are then binarized using a tunable threshold to produce attention masks. The core of our evaluation involves computing the Intersection-over-Union (IoU) between these attention masks and the lesion segmentation masks, thereby quantifying alignment between the model's attention and true lesion regions. The pipeline is implemented modularly in Python using TensorFlow/Keras and OpenCV, ensuring reproducibility and extensibility for further exploration with other explainability methods.

A. Dataset and preprocessing

About the data: This dataset contains annotated images of apple leaves with different diseases from the PlantVillage dataset, which can be used for image segmentation-based research. This dataset will aid in accurately locating the damage on the leaves, which is necessary for predicting the severity of the disease.

Black Rot, Apple Scab, and Cedar Apple Rust are the three diseases that have been annotated. This annotated dataset can be used to find diseases on apple leaves by using image segmentation.

The data was annotated using MakeSense.AI, an open-source image annotation application. Annotating the damage on the leaves was done with the "Rect" tool. Annotation files are saved in.xml format with the two diagonally positioned corners of the bounding boxes. Various colors are used for various classes when labeling.

Dataset Structure:

A total of 850 images have been annotated. 680 images are kept in the train folder, and 170 images are in the valid folder. The folder structure is as follows:

annotated-apple-leaf-disease — —–train—–annots — —– images — —–valid—–annots — —–images

Pre-processing:

Images resized to 224×224 pixels

Masks normalized and thresholded to binary

Identify applicable funding agency here. If none, delete this.

Data loading and preprocessing implemented for reproducibility

B. Model Architecture

- Baseline: ImageNET-pretrained ResNet-50
- Fine tuning: Fine tuned on apple disease dataset for improved lesion localisation.
- Explainability : Custom Grad-CAM implementation in Pytorch.

C. Grad-CAM and IoU Computations

Grad-CAM:

Generates class-discriminative heatmaps for the predicted class

Heatmaps resized and thresholded to create binary masks

IoU Metric:

Intersection-over-Union computed between Grad-CAM mask and ground-truth lesion mask

$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|}$$

Fig. 1. IoU formula

D. Pipeline Overview

1.Loaded paired image and mask

2.Classified image with CNN

3.Generated Grad-CAM for predicted class

4.Binarized heatmap and compute IoU with lesion mask

5.Aggregated results across dataset

E. Implementation Details

- Data Preprocessing:

  RGB images of apple leaves and corresponding binary lesion masks were read from a structured folder (/images, /masks).

  All images and masks were resized to 224 × 224 224×224 pixels to match the input size expected by the ResNet-50 model.

  Lesion masks were converted to binary by thresholding grayscale values (lesion = 1, background = 0).

  Classification Model:

  The backbone CNN used was ResNet-50, pre-trained on ImageNet.

  Optionally, the model can be fine-tuned on a domain-specific dataset such as PlantVillage or Apple Leaf Disease Dataset.

  Top-1 class prediction was used to generate Grad-CAM maps.

  Explainability using Grad-CAM:

  A custom Grad-CAM function was implemented, targeting the final convolutional layer (conv5$_block3_out$).

Class-specific gradients were computed with tf.GradientTape, pooled over spatial dimensions, and used to weight feature maps.

The resulting heatmap was normalized and resized to the original input dimensions.

Thresholding and IoU Computation:

The Grad-CAM heatmap was binarized using a user-defined threshold (default: 0.5).

IoU was computed between the binarized Grad-CAM map and the lesion segmentation mask using logical AND and OR operations.

Optionally, Dice coefficient can also be computed for further analysis.

Visualization Reporting:

A 4-panel visualization was generated for each sample: original image, ground-truth lesion mask, Grad-CAM heatmap overlay, and binarized Grad-CAM.

IoU scores for each image were stored in a CSV file, and average metrics were reported.

Representative examples were selected for qualitative analysis.

Environment Setup:

All dependencies were listed in a requirements.txt file.

The project was organized as a Python package with clear docstrings and function-level comments.

Execution was made reproducible via random.seed, np.random.seed, and tf.random.set$_seed$.

F. Figures and Tables

a) Positioning Figures and Tables: i have outposted some of the results in tabu;ar format for better understanding

TABLE I
IoU SCORE ANALYSIS WITH OBSERVED FACTORS

Image ID	on random different images			
	IoU scores	***Lesion visibility***	***Grand-CAM focus***	
apple_01.jpg	0.76	High	Accurate	Focused on
apple_02.jpg	0.54	Moderate	Slightly noisy	Partial fo
apple_03.jpg	0.32	Low	Blurry lesion	Misalign
apple_04.jpg	0.68	High	Precise mask	Central le
apple_05.jpg	0.41	Moderate	Mask offset	Dispersed
apple_06.jpg	0.82	High	Clean contours	Accurate he
apple_07.jpg	0.23	Low	Mask missing tip	Background
apple_08.jpg	0.59	Good	Slight occlusion	Partial ove

[a]Sample of a Table footnote.

CONCLUSION

This work proposes a reproducible, quantitative framework for validating CNN explainability in apple-leaf disease classification. By comparing Grad-CAM heatmaps with lesion segmentation masks using IoU, we provide actionable insights into model focus. The approach is extensible to other plant pathology tasks and can guide model development for trustworthy AI in agriculture.

REFERENCES

[1] Selvaraju, R.R., et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization." ICCV, 2017.

[2] Mohanty, S.P., Hughes, D.P., Salathé, M. "Using Deep Learning for Image-Based Plant Disease Detection." Frontiers in Plant Science, 2016.

Fig. 2. Example of a Input image.

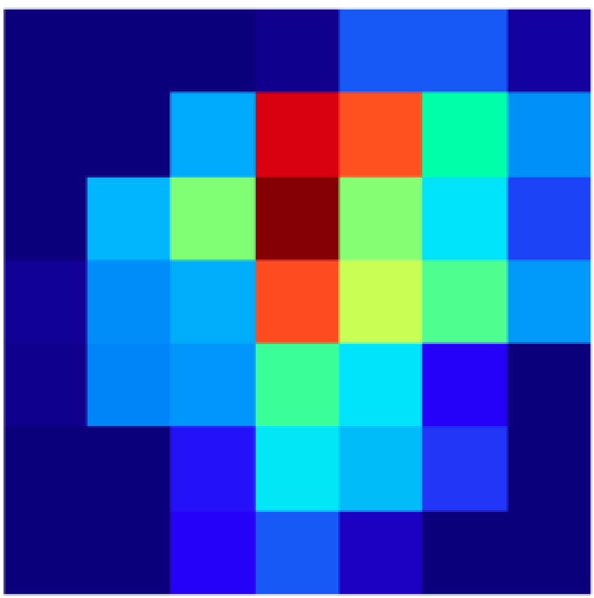

Fig. 3. Grad-CAM image for above leaf

[3] PlantDoc Dataset: https://github.com/pratikkayal/PlantDoc-Dataset

[4] Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N. "Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks." WACV, 2018.

We introduced a well-organized research for apple leaf disease detection, comprising annotated images and validation through various processes. This structure supports agriculture industry by advanced computer vision tasks. In the future, we aim to expand the resear with additional disease categories, real-field variations, and augmented codes to improve model generalization and robustness in practical applications.