Received: xxxxxxxx | Revised: xxxxxxxx | Accepted: xxxxxxx

DOI: xxxxx/xxxxx

ORIGINAL RESEARCH

Integrating Custom CNN Models with SHAP and LIME for Transparent Maize Leaf Disease Classification

Md. Ismiel Hossen Abir[1], Mekhala Mariam Mary[2]

[1] International Standard University

[2] BRAC University

Abstract

Corn is one of the most crucial crops in the worldwide. Sometimes, corn production decreases due to disease, which can affect our food chain and food security. But early detection of the disease is essential for effective food production. In this study, we develop a custom Convolutional Neural Network (CNN) model for corn leaf disease classification with Explainable AI (XAI) techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). For this proposed model, we use four classes of corn leaf conditions, such as Healthy, Blight, Common Rust, and Gray Leaf Spot. Custom CNN model with XAI, we aim to provide actionable insights into the model's decision-making process for AI-driven agricultural solutions. Our CNN model achieved a training accuracy of 99.94%, validation accuracy of 92.71%. A comparative analysis with pre-trained CNN models such as ResNet50, VGG16, MobileNetV2, and DenseNet121 underscores the effectiveness of the custom model in balancing accuracy. We demonstrate SHAP, LIME, and GRAD-CAM to correctly analyze the area of the disease. The results demonstrate the potential of integrating XAI with deep learning to revolutionize disease detection in agriculture.

Keywords: CNN, XAI, Corn leaf diseases

1 | INTRODUCTION

Agriculture plays a crucial role in the global population. Corn is one of the most important crops worldwide, and it is also the most nutritious among all other crops. Corn is the third-largest food source in human civilization. However, the corn production can be threatened by various leaf diseases and making our food chain system insecure and economically imbalanced. But if the leaf disease in the early stage is detected, we can use disease management effectively and ensure the best food production.

The recent advancement of deep learning and computer vision reflects a great effort to automatically detect and classify diseases. Convolutional Neural Network with enhanced interpretability, making the models more useful.

Our contributions can be summarized as:

(i) We develop a CNN-based model for maize leaf disease detection and integrate SHAP and LIME to enhance the interpretability of the model's predictions.

(ii) By applying SHAP and LIME, we provide detailed insights into how the model arrives at its decisions, enabling users to better understand and trust the AI system.

(iii) We conduct a comprehensive comparative analysis between custom CNN and pre-trained CNN models on the Corn or Maize Leaf Disease Dataset, evaluating their performance in terms of accuracy.

2 | RELATED WORK

Ahad et al. [1] conducted a study comparing six CNN-based deep-learning architectures, and they found that ensemble models significantly improved accuracy in detecting rice leaf diseases. Masood et al. [2] proposed the MaizeNet model, a deep learning framework based on Faster-RCNN with ResNet-50 and spatial-channel attention for effective localization and classification of maize plant leaf diseases. Sagar et al. [3] conducted a comprehensive survey on leaf-based plant disease detection techniques using deep learning models and explored the use of Explainable AI (XAI) to make these models more transparent. Zhengjie Ji et al. [4]

proposed a lightweight convolutional neural network, ICS-ResNet, based on ResNet50, to improve the accuracy of maize leaf disease classification. Baliyan et al. [5] propose a deep learning- based CNN model for multi-classification of corn gray leaf spot (CGLS) disease into five severity levels, achieving a high detection accuracy of 95.33 percent. Goldwasser et al. [6] presented methods for ensuring stable feature rankings using SHAP and LIME by addressing the instability inherent in these popular attribution techniques. Kundu et al. [7] proposed a deep learning-based framework for automatic maize disease detection, severity prediction, and crop loss estimation, achieving a high accuracy of 98.50 percent. Paymode et al. [8] work focuses on using VGG-based CNNs for early detection and classification of leaf diseases in crops, achieving high accuracy rates of 98.40 percent for grapes and 95.71 percent for tomatoes.

In our research, we uniquely integrated custom CNN models with explainable AI techniques like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to enhance the transparency and interpretability of maize leaf disease classification. We provided clear insights into the decision-making process of the model, making our approach more reliable and comprehensible for end-users and researchers alike.

3 | METHODOLOGIES

3.1 | Data Preprocessing and Overview

The Corn leaf disease dataset [11] is a comprehensive dataset of images designed with four classes that contain both healthy and disease-affected corn leaves. These images are crucial for training deep learning models to detect leaf disease, which is an essential task for enhancing agricultural productivity and ensuring food security. According to Figures 1 and 2, we use a set of images from all four classes in the dataset: Healthy, Blight, Common Rust, and Gray Leaf Spot. Each class is represented with distinct leaf appearances, showing the variations in disease symptoms.

Blight: A condition where leaves exhibit irregular, brown spots.

Common Rust: Characterized by orange or reddish pustules on the leaf surface.

Gray Leaf Spot: Identified by small, rectangular lesions on the leaves.

Healthy: Images of corn leaves with no visible disease symptoms.

The number of images in each of the classes is divided into three sections: training (3254), testing (409), and validation (405) images. We use a total of 80% images for training, 20% of images, and 20% images for validation purpose. The total number of images is 4068.

Preprocessing is an essential step in preparing the maize leaf images for training the CNN model. In this study, several preprocessing techniques were applied to ensure that the data was suitable for training the model. In the preprocessing steps included image resizing, data augmentation techniques, and normalization were included. All images were resized to a uniform dimension to match the input requirements of the CNN model; we use 150X150X3. Normalization was applied to scale the pixel values between 0 and 1. Additionally, data augmentation techniques such as rescaling, rotation, horizontal and vertical flipping, zooming, shear range, width and height shift, and brightness adjustment were employed. These augmentations artificially increased the diversity of the dataset, allowing the model to learn from a broader range of variations, which is particularly important for improving the model's ability to generalize across different environmental conditions and disease manifestations.

Figure 1: Sample Dataset Visualization

3.2 | Model Architecture

According to Figure 2, we have developed a Convolutional Neural Network (CNN) for image classification into four categories. We use five convolutional blocks in the model architecture. First 4 convolutional blocks, we use two convolutional layers, and we use one convolutional layer in the last one. Each of the convolutional layers has 64 filters, a kernel size of (3, 3), and ReLU activation. Each block is followed by a max-pooling layer and batch normalization to enhance training stability and reduce overfitting. Dropout layers are strategically placed after certain convolutional and fully connected layers with dropout rates between 0.25 and 0.3 to prevent overfitting. Additionally, L2 regularization (0.0001) is applied to the dense layers. The network concludes with a fully connected layer with 128 neurons (ReLU activation) and a softmax output layer with four neurons.

Table 2 shows the figure 3 model summary. Our architecture effectively captures image features while maintaining model efficiency, ensuring performance across varying image categories. After that, for the compiled purpose, we use Adamax optimizer with an initial learning rate of 0.001 and categorical cross-entropy loss to handle the multi-class classification task. To enhance training efficiency and prevent overfitting, we also use several callback functions such as Learning Rate Scheduler, Reduce LROnPlateau, and Early Stopping. A custom learning rate scheduler was implemented, which exponentially decays the learning rate over epochs. The Reduce LROnPlateau callback monitors the validation loss and reduces the learning rate by a factor of 0.2 if the validation loss does not improve for five consecutive epochs, with a minimum learning rate threshold of 0.0001. And, we use a custom callback function, which will early stop when the accuracy or validation accuracy reaches 99. The model was trained for 100 epochs with a batch size of 32 using these callbacks, ensuring efficient learning across the train- ing and validation datasets.

DenseNet121. These models are fine-tuned on the maize leaf disease dataset to feature extraction capabilities. The choice of multiple architectures aims to compare their performance and adaptability in recognizing distinct leaf disease patterns.

For explainable AI, we use LIME and SHAP. LIME is used to interpret the model's predictions by approximating the local decision boundary around each prediction. SHAP values are computed to quantify the contribution of each pixel in the input image to the model's predictions. We also use Grad-CAM, which is used to visualize the areas of the images that are most influential in the model's decision-making process.

4 | EXPERIMENTAL RESULTS AND ANALYSIS

The results of the experiments provided into three parts those are CNN models accuracy and report, explainable AI, and compare with per-trained model. Validation accuracy of the proposed CNN model is 92.71, test accuracy is 91.20 and training accuracy of the model is 99.94.

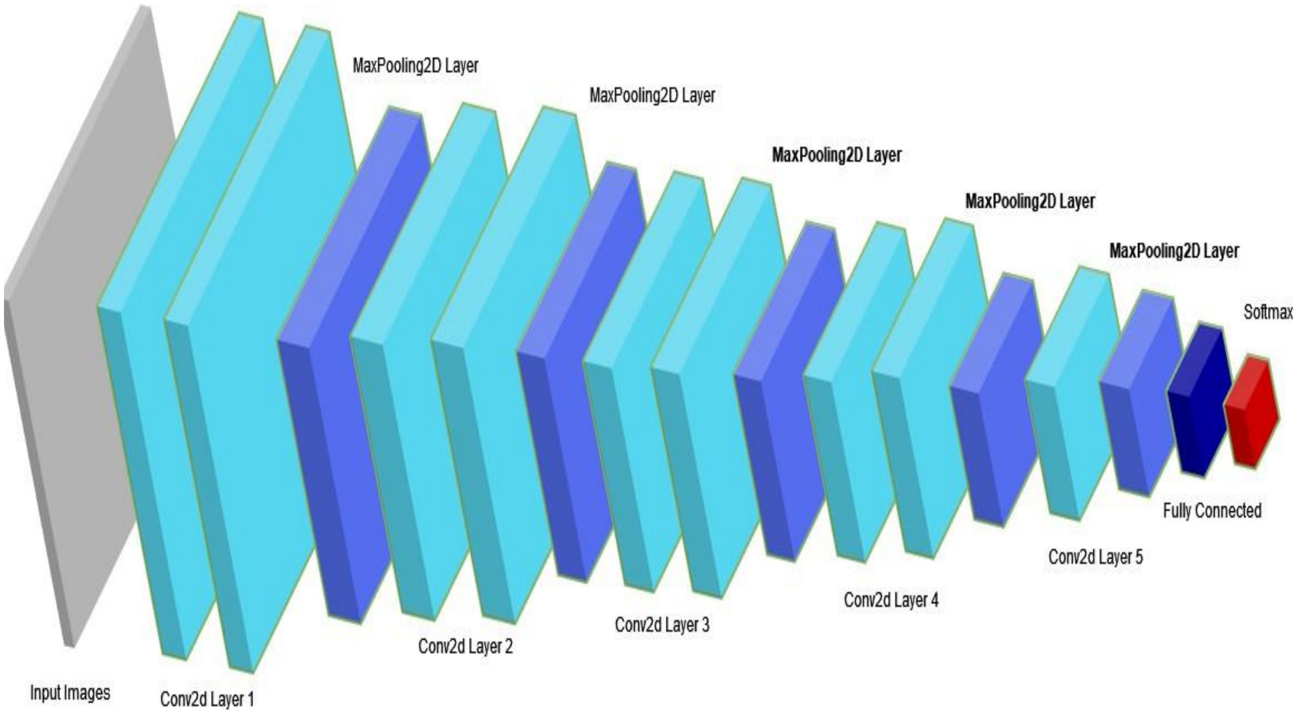

Figure 2: CNN Model Architecture

The classification report for our custom CNN model on the maize leaf disease dataset is presented in Table 1. The report includes the precision, recall, F1-score, and support for each class: Blight, Common Rust, Gray Leaf Spot, and Healthy. The model has achieved an overall accuracy of 91 percent for Common Rust, and Healthy classes had the highest performance with F1-scores of 0.94 and 0.97. The macro and weighted averages highlight the balanced performance across all classes.

The confusion matrix in Figure 4 reflects the performance of the custom CNN model across the four classes: Blight, Common Rust, Gray Leaf Spot, and Healthy. The model

Table 1: Heading of the table

	Precision	**Recall**	**F1-Score**
Blast	0.87	0.86	0.87
Common Rust	0.93	0.94	0.96
Gray leaf disease	0.82	0.78	0.80
Healthy	0.97	0.97	0.97
Accuracy			0.92.71

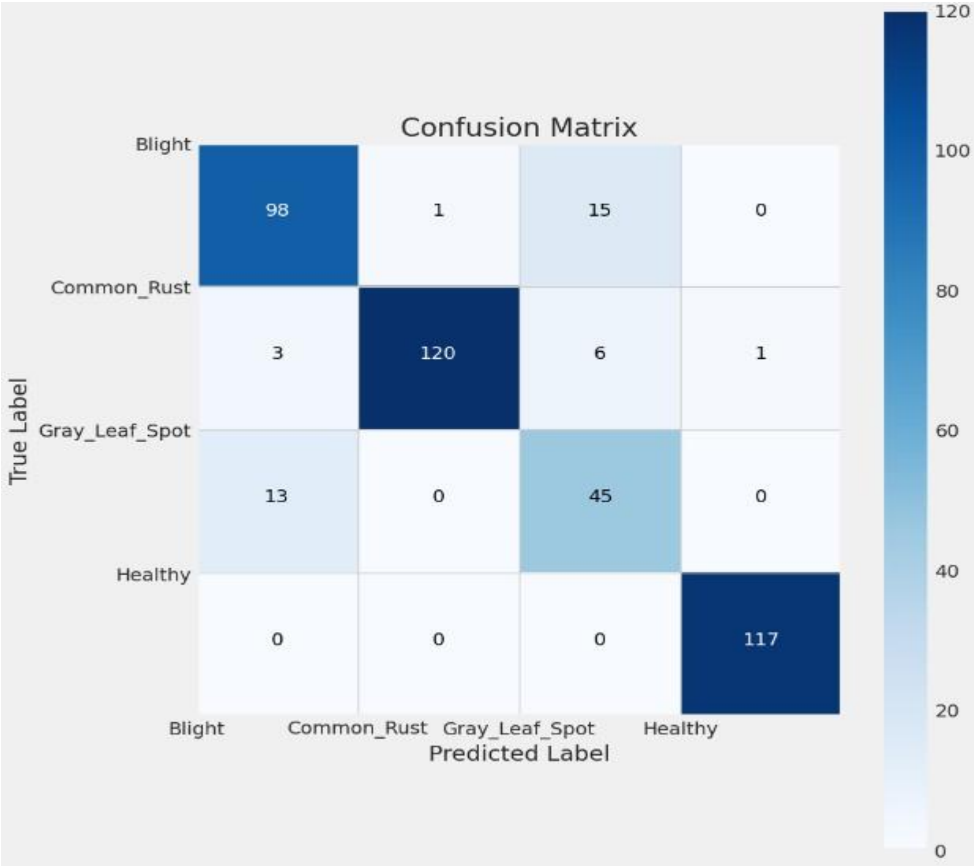

Figure 3: Confusion Matrix

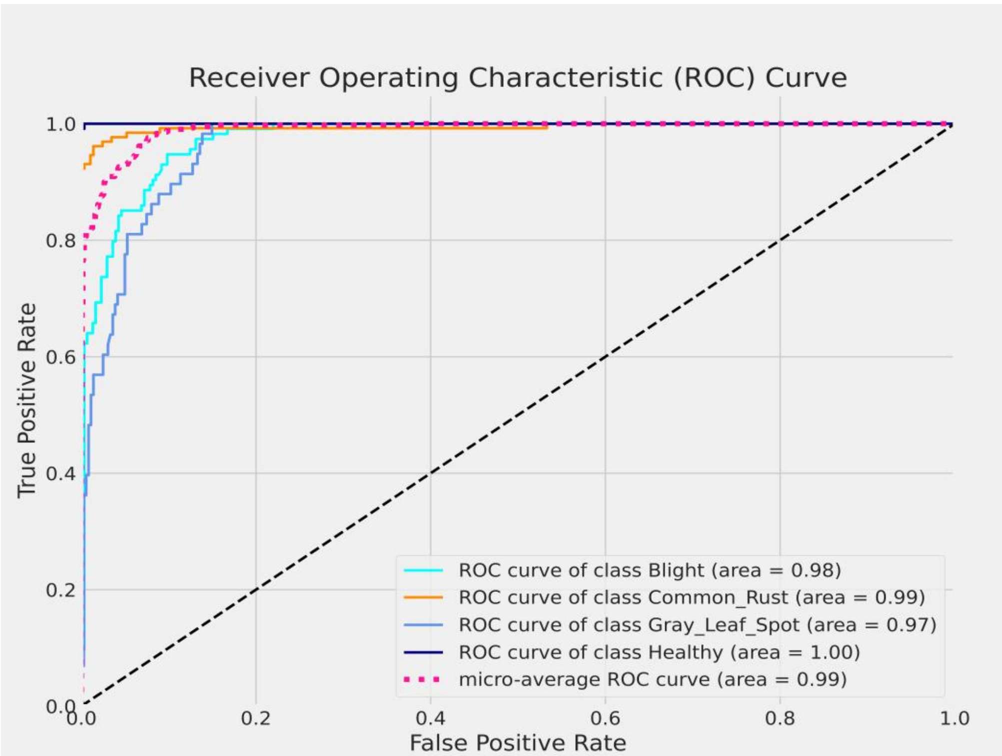

Figure 4: ROC Curve and AUC Score

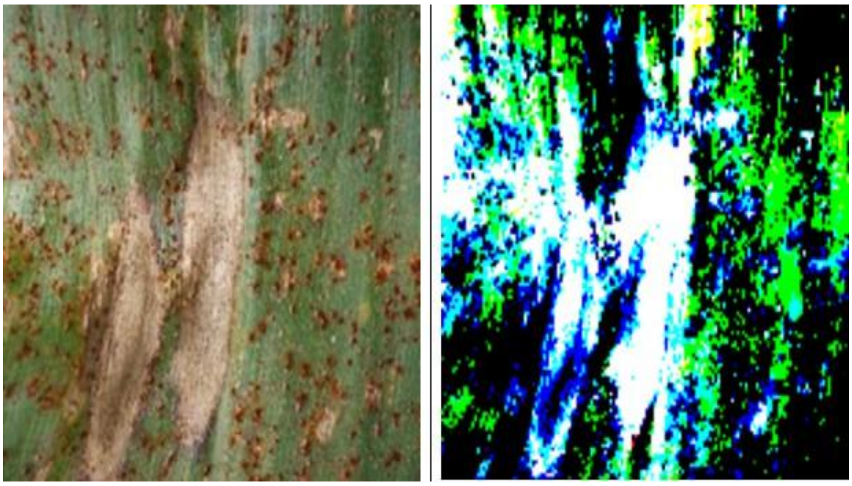

Figure 5: Original and LIME for Blast

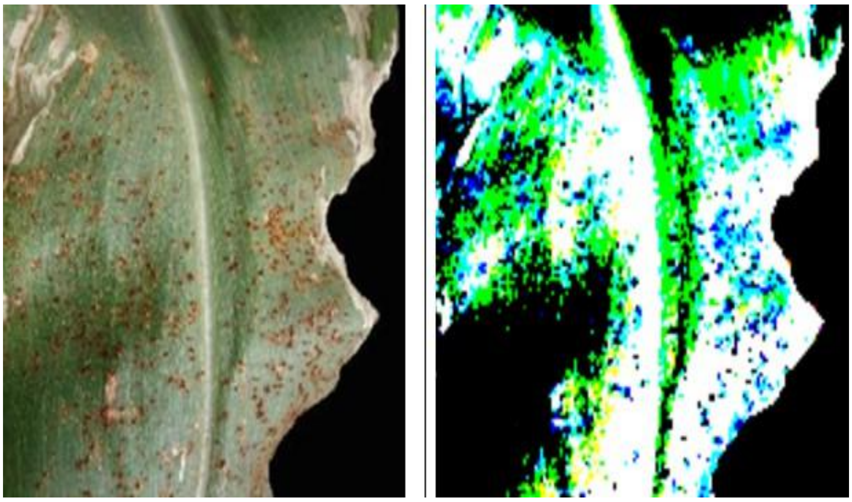

Figure 6*: Original and LIME for Common Rust*

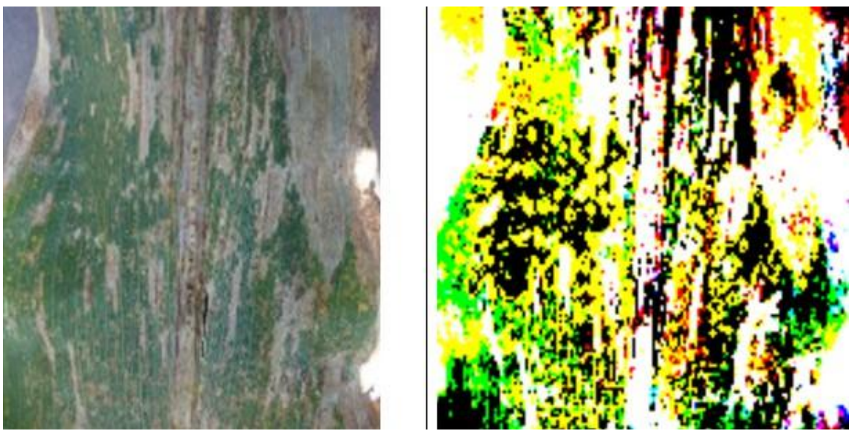

Figure *7: Original and LIME for Gray Leaf Spot*

The classification report for our custom CNN model on the maize leaf disease dataset is presented in Table 1. The report includes the precision, recall, F1-score, and support for each class: Blight, Common Rust, Gray Leaf Spot, and Healthy. The model has achieved an overall accuracy of 91 percent for Common Rust, and Healthy classes had the highest performance with F1-scores of 0.94 and 0.97. The macro and weighted averages highlight the balanced performance across all classes.

The confusion matrix in Figure 3 reflects the performance of the custom CNN model across the four classes: Blight, Common Rust, Gray Leaf Spot, and Healthy. The model shows high accuracy in predicting Common Rust and Healthy classes, with minimal misclassifications.

According to the figure 4, the model demonstrates strong performance in classifying plant diseases. The micro-average ROC curve, with an area under the curve (AUC) of 0.99, indicates excellent overall discrimination between healthy and diseased plants. Individual classes also show high AUC values, with Common Rust achieving the best performance at 0.99. Blight and Gray Leaf Spot exhibit similarly good results, while Healthy plants are perfectly classified with an AUC of 1.00.

In explainable AI, we use LIME, SHAP, and for visualization, we use the GradCAM method. In LIME (Local Interpretable Model-agnostic Explanations) visualizations, different colors help users understand which parts of the image are most influential in the model's decision-making process. In the Figure 5, it represents white and green colors indicate positive identification of the disease in the blight class example image. On the other hand, dark or blue reflects that those areas are not affected by the disease. Figures 6 and 7, it show that white, green, and yellow color regions are disease spot, and other dark and blue areas are not affected.

Here, figure 8, 9 each of the figure reflects the SHAP values each of the classes. Each of the class the first image is original image. This is the input image that the model is making predictions on. In this case, it appears to be an image of a leaf with some spots on it. Output 0, Output 1, Output 2, Output 3 are the four possible output categories of the model. The SHAP values will explain how the features of the input image contribute to the predicted probability of each category. At the bottom of each figure there are a scale called SHAP value. This is the color-coded bar. It represents the range of possible SHAP values, from -2 to 2. Blue SHAP values indicate that a feature decreases the probability of a category, while red SHAP values indicate that a feature increases the probability of a category. Red colors usually indi- cate pixels or regions that have a positive impact on the predicted class. On the other hand, Blue colors typically represent pixels or regions that have a negative impact on the predicted class.

Another term, Grad-CAM (Gradient-weighted Class Activation Mapping) is a method for visualizing deep learning model predictions, especially for convolutional neural networks (CNNs) in image classification tasks. Grad-CAM visualizations use a heatmap to indicate the areas of an image that are most relevant to a model's decision for a specific class. The heatmap is overlaid on the original image for easy interpretation. According to the figure 10, 11 it represent the GRAD-CAM of the corn leaf disease classes. Each of the figure have 3 images, first one for original, second one for heatmap and the last one for GradCAM overlay. The heatmap highlights the regions of the image that are most important for the model's prediction. Redder areas indicate higher importance, while bluer areas indicate lower importance. And grad-cam overlay shows original image with the heatmap overlaid on top of it. This visualization helps to identify which specific regions of the leaf are contributing most to the model's prediction.

After that, we use 4 pre-trained models to compare and observe our custom model's result. So, according to Table 2, we found that our custom CNN model gives higher accuracy than the other 4 pre-trained models.

Table 2: Model Comparison

Model Name	**Accuracy Score**
Resnet50	92.12
VGG16	87.34
MobileNetV2	87.77
DenseNet121	86.36
Custom CNN Model	92.71

5 | CONCLUSION

In conclusion this study, we developed and evaluated a custom Convolutional Neural Network (CNN) model for maize leaf disease detection, using state- of-the-art Explainable AI (XAI) techniques, namely SHAP and LIME. Our approach aimed to not only achieve high accuracy in disease classification but also enhance the interpretability of the model's predictions. The custom CNN model demonstrated robust performance with a validation accuracy of 92.71 percent, outperforming several established pre-trained models such as ResNet50, VGG16, Mo- bileNetV2, and DenseNet121, which recorded ac- curacies of 92.12 percent, 87.35 percent, 83.77 per- cent, and 86.36 percent, respectively. The model's effectiveness was further confirmed through comprehensive evaluation metrics, including precision, recall, F1-score, and AUC, which reflected high performance across all disease categories. Future work can explore further optimization of the model, potentially incorporating additional environmental variables and expanding the dataset to improve generalization. Additionally, extending the integration of XAI techniques could provide even deeper insights into the model's behavior, further supporting stakeholders in making informed agricultural decisions.

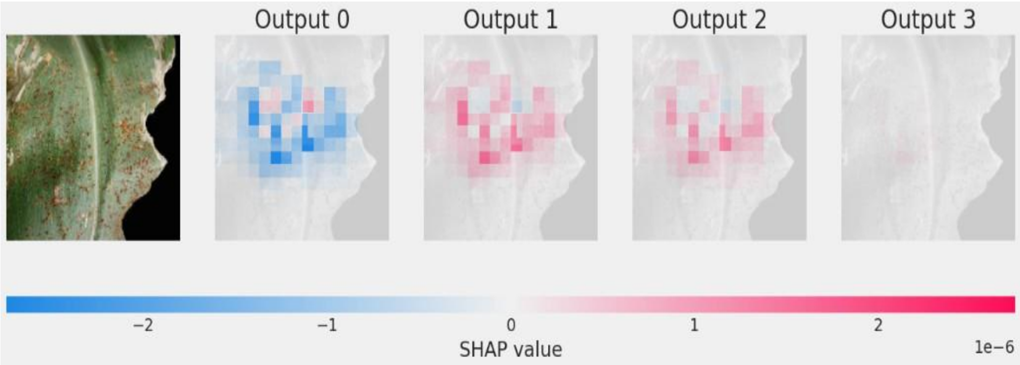

Figure 8: SHAP for Common Rust

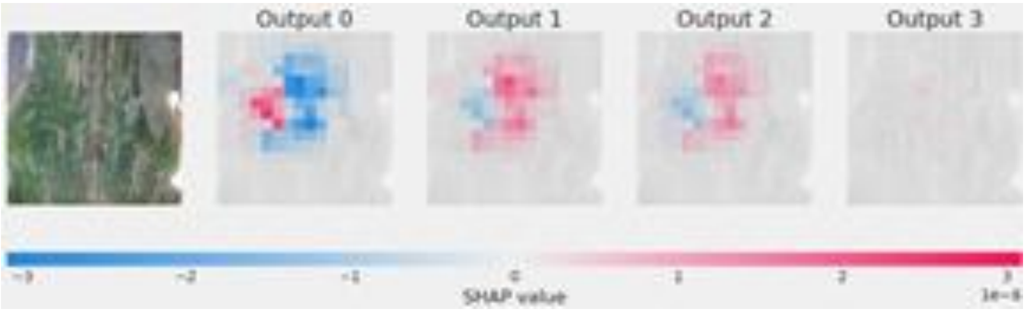

Figure 9: SHAP for Gray Leaf Spot

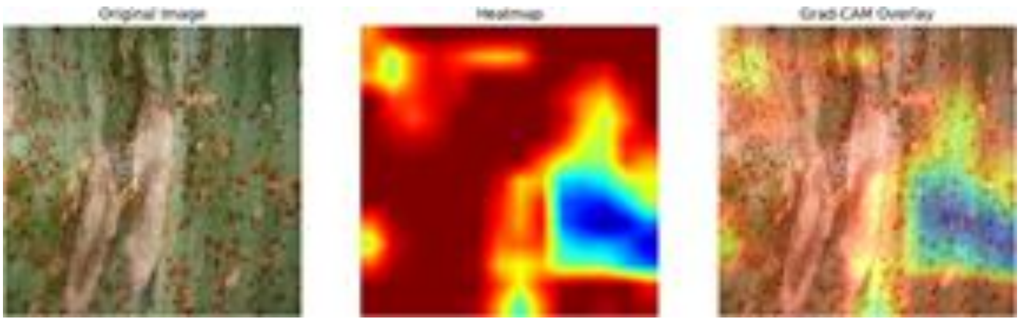

Figure 9: GRAD-CAM for Blight Class

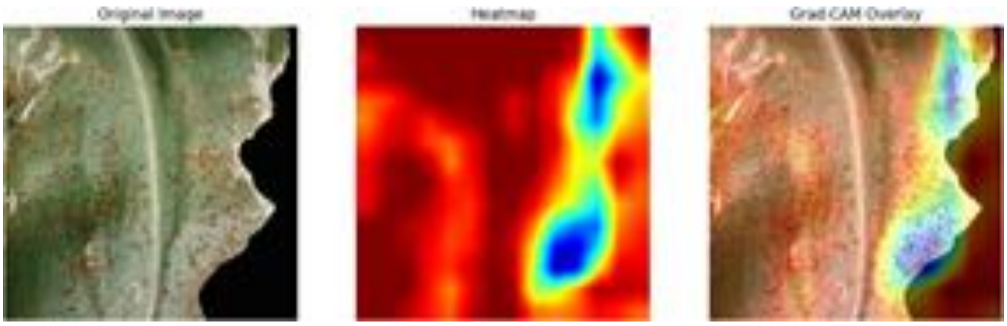

Figure 9: GRAD-CAM for Common Rust Class

REFERENCES

[1] M. T. Ahad, Y. Li, B. Song, and T. Bhuiyan, “Comparison of CNN-based deep learning architectures for rice diseases classification,” Artificial Intelligence in Agriculture, vol. 9, pp. 22–35, Jul. 2023, doi: 10.1016/j.aiia.2023.07.001.

[2] Masood, M., Nawaz, M., Nazir, T., Javed, A., Alkanhel, R., Elmannai, H., Dhahbi, S., Bourouis, S. (2023). MaizeNet: A deep learning approach for effective recognition of maize plant leaf diseases. IEEE Access, 11, 52862-52872. https://doi.org/10.1109/ACCESS.2023.3280260

[3] S. Sagar, M. Javed, and D. S. Doermann, ”Leaf-Based Plant Disease Detection and Explain- able AI,” arXiv preprint arXiv:2404.16833v1, 2023.

[4] Z. Ji, S. Bao, M. Chen, and L. Wei, ”ICS- ResNet: A Lightweight Network for Maize Leaf Disease Classification,” The College of Informa- tion Science and Technology, Gansu Agricultural University, Lanzhou 730070, China, 2024.

[5] A. Baliyan, V. Kukreja, V. Salonki, and K. S. Kaswan, ”Detection of Corn Gray Leaf Spot Severity Levels Using Deep Learning Approach,” in Proc. 9th Int. Conf. on Reliability, Infocom Technologies and Optimization (Trends and Fu- ture Directions), Amity University, Noida, India, Sep. 2021, pp. 1-6, doi: 10.1109/ICRITO51393.2021.9596540.

[6] J. Goldwasser and G. Hooker, ”Provably Stable Rankings with SHAP and LIME,” Depart- ment of Statistics, UC Berkeley and Department of Statistics and Data Science, University of Penn- sylvania, 2024.

[7] N. Kundu, G. Rani, V. S. Dhaka, K. Gupta, S. C. Nayaka, E. Vocaturo, and E. Zumpano, ”Dis- ease detection, severity prediction, and crop loss estimation in MaizeCrop using deep learning,” Com- puters and Electronics in Agriculture, vol. 207, pp. 107658, Nov. 2022.

[8] A. S. Paymode and V. B. Malode, ”Trans- fer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG,” Artificial Intelligence in Agriculture, vol. 6, pp. 23–33, 2022.

[9] Singh D, Jain N, Jain P, Kayal P, Kumawat S, Batra N. PlantDoc: a dataset for visual plant disease detection. InProceedings of the 7th ACM IKDD CoDS and 25th COMAD 2020 Jan 5 (pp. 249-253).

[10] J, ARUN PANDIAN; GOPAL, GEETHARAMANI (2019), “Data for: Identification of Plant Leaf Diseases Using a 9-layer Deep Convolutional Neural Network”, Mendeley Data, V1, doi: 10.17632/tywbtsjrjv.1

[11] S. Ghose, ”Corn or Maize Leaf Disease Dataset,” Kaggle, 2020. [Online]. Available: https://www.kaggle.com/datasets/ smaranjitghose/corn-or-maize-leaf-disease-dataset. [Accessed: Aug. 15, 2024].