Identifying potato diseases with explainable AI

1st Salwa Tamkin
Dept of CSE
BRAC University
Dhaka, Bangladesh
salwatamkin4@gmail.com

2nd Md. Ashikur Rahman
Dept of CSE
BRAC University
Dhaka, Bangladesh
md.ashikur.rahman6@g.bracu.ac.bd

***Abstract*—Crop diseases inflict significant economic damage on a global scale. Their primary impact lies in drastically reducing crop yields, which translates to higher food prices for consumers and potential food shortages, particularly affecting vulnerable populations. These widespread consequences highlight the critical need for effective disease management strategies to protect agricultural productivity. This study addresses this challenge by employing Convolutional Neural Networks (CNNs) to accurately identify diseased leaves. Furthermore, Explainable Artificial Intelligence (XAI) techniques are implemented to understand the decision-making processes within the models. Two distinct CNN models, VGG19 and Xception, were utilized for image classification, achieving accuracy ranging from 78.47% to 83.54%. The integration of XAI provides valuable insights into the models' reasoning behind their classifications, empowering farmers to understand disease presence and take appropriate precautions to mitigate potential widespread crop failures and economic crises.**

I. Introduction

Global food security is severely endangered by plant diseases. Fungi, bacteria, and even certain parasitic plants can cause these diseases, which can severely damage crops, limiting yields and compromising the availability of food. Global potato production is impacted by yield losses, which range on average from 8.1% to 21%. Smallholder farmers are severely affected; reports suggest that production losses for potatoes in South America can be as high as 50% [1]. The traditional disease detection techniques can be labor-intensive, time-consuming, and require specialized people. However, the development of machine learning (ML) provides an effective method for image-based plant disease identification, much faster than human experts, allowing timely and rapid disease detection.

This study focuses on early identification of potato diseases, leveraging machine learning algorithms to distinguish between healthy and diseased plants. By incorporating explainable artificial intelligence (XAI) tools, this research aims to enhance the model's interpretability beyond the traditional plant disease identification approach. The desire to find explainability in AI models is increasing since XAI can help users establish credibility in the model by offering visual explanations of the model's decision-making process and reasoning behind the model's predictions. The term "black boxes" is frequently applied to CNNs, which raises concerns regarding bias, accountability, and trust. Our research attempts to address these issues by shedding light on the inner workings of the CNN architecture. This will enable researchers, developers, and users to make more informed decisions about which image elements are used to classify the diseases in the image data. Our goal is to close the gap between model complexity and human understanding by combining explainable AI with Machine Learning (ML) and Convolutional Neural Networks (CNNs).

II. Background Study

This paper [2] uses optical images to investigate the potential of deep learning models for automated detection of early and late blight diseases in potato leaves. The PlantVillage dataset was used to train four deep learning architectures: VGG16, VGG19, MobileNet, and ResNet50. Among the models, VGG16 proved to be the most successful, attaining an accuracy rate of 92.69%. The study used fine-tuning, modifying the model's parameters in response to particular data, to further enhance performance. The optimized VGG16 model distinguished between potato leaves with late or early blight with an impressive 97.89% accuracy rate.

Additionally, Computer vision, utilizing image processing techniques, offers a promising solution for rapid and efficient potato disease identification. This study [3] explores the field of potato diseases and how AI and machine learning may be used to fight them. It provides a thorough overview of the primary potato diseases with an emphasis on using computer vision techniques to identify them. The assessment concludes that the three most frequent threats to potato crops are bacterial wilt, early blight, and late blight after conducting a thorough analysis of 39 relevant research papers. In addition, it draws attention to the expanding trend of using deep learning algorithms rather than more conventional machine learning techniques for disease identification.

Another research work [4] suggests an automated method that combines machine learning and image processing. This method uses a Support Vector Machine (SVM) classifier combined with image segmentation to automatically diagnose diseases in potato plants using the publicly accessible PlantVillage dataset. This method applies multi-class SVM for image segmentation, yielding an accurate 95% on more than 300 images, and provides a user-friendly, automated disease diagnostic tool.

Additionally, a study [5] employs a CNN model and a deep learning technique to scan potato leaf images in order

to classify diseases. In order to do this, a dataset of 10,000 images was put together from several places, such as Google and actual potato fields. 2,152 photos from Kaggle were added to the dataset to further enhance it. Three classes are included in this diverse dataset: potato early blight, potato late blight, and healthy potato leaves. According to the study, when trained with 40 epochs on the 10,000-image dataset, the CNN model obtained the greatest accuracy (100%) of any model tested. Furthermore, accuracy levels of 99.97% and 99.98% were obtained by training with 30 and 50 epochs, respectively.

While the research focuses on potato diseases, these advanced techniques are applicable to a wider range of crops, including corn, apple, tomato, and many others. This study [6] proposes a multi-model method using three CNN architectures: EfficientNetV2L, MobileNetV2, and ResNet152V2, which significantly advances the field of plant disease identification. This method successfully identifies 38 illnesses in 14 different plant species, demonstrating its robustness and adaptability to various plant disease scenarios. EfficientNetV2L is the best performer, according to performance evaluation, with an amazing accuracy of 99.63%. Additionally, incorporating Explainable AI (XAI) frameworks like LIME enhances the transparency of the models by providing interpretable explanations for their predictions. This transparency builds trust and confidence in the model's accuracy and decision-making process.

III. Methodology

The primary objective of this study is to classify diseased plants from images using explainable artificial intelligence techniques, which will enhance the classification model's interpretability and transparency. The study uses explainable AI that expands beyond prediction accuracy and offers comprehensible reasoning for the classification predictions generated by the algorithm. The suggested method requires gathering a dataset. Our research methodology mainly examines the classification accuracy of two pre-trained models, VGG19 and Xception, along with data augmentation. We further incorporate the interpretability of the two pre-trained models using Grad-CAM (Gradient-weighted Class Activation Mapping) to consistently identify images of healthy and unhealthy potato plant leaves, thereby justifying the comprehension of their decision-making process.

A. Dataset Preparation and Augmentation

Our research on comparing plant disease detection methods utilizes the Bangladeshi Crops Disease Dataset [7]. This image-based dataset contains images of four crops (corn, rice, wheat, and potato) in various disease states. With a total of 13,024 images categorized into 14 classes, we have specifically selected two classes for our analysis: Potato Late Blight (diseased) and Potato Healthy with the intention to identify late blight disease in potatoes with deep learning models. The selected dataset was comprised of total 5137 images, where the number of healthy potato leaf images was 2006 and the rest of the 3131 images belonged to the late blight potato class. This taken dataset was categorized into a 60:40 ratio for training and test set, respectively, as shown in Table I. This focused selection allows for a more in-depth comparison of the methods. illustrates representative samples from each crop type included in this dataset.

TABLE I
Details of the Bangladeshi Crops Disease Dataset

Classes	Number of Images	Training Set	Test Set	Repository
Potato Healthy	2,006	1,204	802	Kaggle
Potato Late Blight	3,131	1,880	1,251	

An extensive procedure of data augmentation and preprocessing was applied to the image dataset in order to optimize the efficacy of the disease detection models. The purpose of this multi-step process was to induce differences in the images and artificially increase the size of the dataset. By reducing the possibility of the model overfitting to a limited dataset, this approach helped the model become more accurate in handling diverse image scenarios.

This extensive data pre-processing and augmentation procedure greatly enhanced the quantity and quality of the dataset, establishing it for the CNN model. Multiple approaches were used, such as:

1) **Scaling and Normalization:** Images were resized to match the input requirements of the neural networks and normalized to a specific range, ensuring compatibility within the model.
2) **Cropping:** Uninformative areas were removed through cropping, ensuring consistency in image size. Maintaining consistency is vital for making efficient use of pre-trained CNN models.
3) **Rotation and Flipping:** Images were rotated, flipped, and slightly shifted to introduce variations in their orientation and position.
4) **Color augmentation:** The contrast, saturation, and brightness of the image were adjusted to increase the overall visual diversity.

Therefore, every image was in .jpg format and compressed to a consistent 224x224 pixel size, resulting in a standardized dataset that can effectively train CNN networks.

B. Comparative Analysis of CNN Architectures for Potato Plant Disease Classification

Convolutional Neural Networks (CNNs) were utilized in our research to address image classification using deep learning techniques. Consequently, the machines were able to efficiently analyze and extract data from the images. In our study, we implement two different CNN architectures and the Grad-CAM framework for Explainable Artificial Intelligence.

VGG19: VGG19 [8] describes a class of deep Convolutional Neural Network (CNN) architectures created by Oxford University's Visual Geometry Group. This deep structure allows them to learn complex features from images, making them highly effective for image classification and object recognition. One such variation in the VGG family is VGG19.

It has 19 convolutional layers total: 16 convolutional layers, 5 max pooling layers, and 3 fully connected layers. These convolutional layers are organized into blocks, with several convolutional layers with small filter sizes (3x3) stacked on top of one another in each block. In order to minimize the dimensionality of the data and prevent overfitting, max pooling layers are also included in between these convolutional blocks whereas fully-connected layers are responsible for determining the final classification. Initially, it was trained using the massive ImageNet dataset, which had over a million images divided into 1000 classes[8]. VGG19 is able to develop detailed feature extractions through this pre-training that can be fine-tuned for various image recognition tasks.

Xception: Xception [9] is a deep learning architecture belonging to the family of Convolutional Neural Networks (CNNs). Developed by Google researchers, it has a notable depth of 71 layers, which is significantly deeper than VGG19[9]. Xception expands upon the ideas of another widespread CNN architecture, the Inception V3 network. It makes use of approaches such as depthwise separable convolutions, which employ two different layers rather than one: a pointwise convolution for channel-wise combination and a depthwise convolution for feature extraction. This method preserves feature extraction capabilities while assisting in the reduction of the number of parameters. The ImageNet dataset, which has over a million images classified into 1000 object classes, and the JFT dataset, which has over 350 million images with high resolution labeled with labels from a set of 17,000 classifications, are typically used to train the Xception network. On both the ImageNet and JFT datasets, Xception exhibits marginal improvements in classification performance. Xception can acquire detailed feature extractions through pre-training, which it can then adjust for different image recognition tasks.

C. Employing Grad-CAM for Visual Explanation of Plant Disease Classification in CNN

Gradient-weighted Class Activation Mapping, or Grad-CAM for short, is a method used in Explainable Artificial Intelligence (XAI) to illustrate a Convolutional Neural Network's (CNN) attention during a classification decision.

CNNs frequently operate like "black boxes," making it challenging to comprehend how they determine their outputs. Grad-CAM computes the gradients of the final class score relative to the convolutional layers' activation maps. A weighted sum of the activation maps from a user-selected convolutional layer is then produced using the computed gradients. Usually, this layer is selected based on the feature extraction. After upsampling the weighted sum to the input image's size, a "class activation map" is produced that shows the areas of the image that the model focused on the most during classification tasks. Although both VGG19 and Xception are strong CNN architectures, it can be challenging to determine which particular areas of an image have the greatest influence on the classification result due to their depth [10]. Grad-CAM facilitates this visualization by producing a heatmap, or class activation map, which highlights the regions of the picture that the model concentrated on during prediction. Additionally, Grad-CAM's visual explanations help boost stakeholders' confidence in the model's decision-making process, particularly those who might be unfamiliar with the technicalities of deep learning.

By applying Grad-CAM to both VGG19 and Xception models, we were able to generate heatmaps that highlight the important regions of the images for classification. This helps us verify whether the models are focusing on the right areas and provides insights into their classification decisions.

IV. Results and Discussion

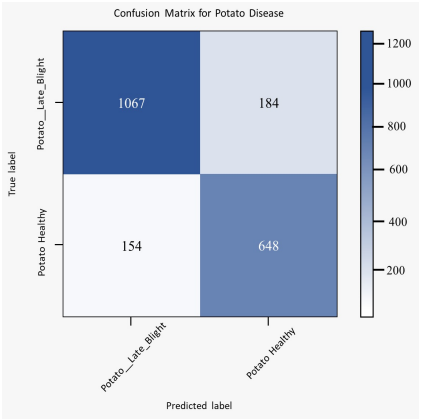

Fig. 1. Confusion matrix for VGG19 model

The confusion matrix in Fig. 1 highlights the VGG-19 model's effectiveness in detecting potato leaf disease. True positive samples totaled 1067, correctly identifying leaves with late blight. However, 184 samples were false negatives, and 154 were false positives, indicating the model misclassified some leaves. Meanwhile, 648 samples were true negatives, confirming healthy potato leaves on both true and predicted labels.

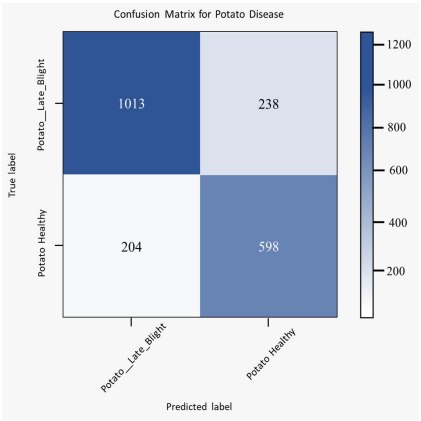

Fig. 2. Confusion matrix for Xception model

In addition, the confusion matrix of the Xception model has been shown here in Fig. 2 where imperative performance of

this model had been also weighed based on true and predicted label. 1013 potato leaf samples are true positives since these are actually diseased on the true label and also been predicted to be affected with potato late blight disease by this model. And the remaining 238 diseased leaves were found to be misinterpreted and can be classified as false negatives. Besides, 598 had been identified correctly on both true and predicted label as healthy leaves and said to be true negatives whereas the rest of the 204 samples were termed as the false positives.

A. Performance Measurement analysis

The two CNN model's evaluation measurements involved in this study had also been analyzed throughout the classification process including accuracy, precession, recall, and F1-score, that have been showed in Table 2.

TABLE II
PERFORMANCE METRICS OF CNN MODELS

CNN Models	Accuracy	Precision	Recall	F1-score
VGG-19	83.54%	87.39%	85.29%	86.33%
Xception Model	78.47%	83.23%	80.98%	82.01%

Table 2 demonstrates the CNN model's performance metrics where VGG19 had served better performance measurements with a promising overall accuracy of 83.54 % over 2053 test potato leaf images taken for this identification process compared to the other Xception model's accuracy of only 78.47%. Even in terms of precision, recall and F1-score, VGG19 outperformed the other model with the precision score of 87.39%, recall of 85.29% and relatively higher 86.33% F1-score. But in case of both CNN models, accuracy is found to be slightly lower compared to the other three performance metrics including precision, recall and F1-score.

In such case, the slightly lower accuracy indicates the models might be struggling with correctly classifying some leaf samples, even though it provided good precision and recall for specific predictions which indicates the models are good at precisely identifying the positive cases (majority class) more effectively. This implies a relative bias towards the majority class (positive cases) since there is a class imbalance problem in this study. The overall dataset was divided into 60:40 ratio respectively for training and test set and the ratio of majority class (positive cases) to minority class (negative cases) was 3:2 which means positive cases are dominant in this case. And so, the engaged models are seemed to prioritize the majority class, resulting into slightly higher precision, recall and F1-score so far. Another possibility can be the borderline classes that could be misclassified. The model might struggle with samples that fall near the decision boundary between classes, leading up to a slight drop in accuracy compared to the more clear-cut positive or negative predictions but still managed to achieve high precision and recall for the classes it confidently predicts. Despite such disparity in performance metrics, the VGG19 model seemed to have noteworthy potential in terms of identifying potato leaves that are affected with late blight disease. By engaging better data augmentation techniques and assigning relatively higher weights to the minority class, this model has the potential to outscore the conventional CNN models and provide a neat balance in performance measurements for the identification of crops disease worldwide.

B. Interpretation of the PrivacyPreserveNet on given datasets

Since model's interpretability deems to add up a cabalistic insight on the better understanding of its decision making, GRADCAM algorithm had been applied in our engaged models to map the salient regions of the target image that is vigorously held responsible for addressing that sample image to a particular class. Overall, figure-3 shows both the histogram of the input image and it's corresponding GRADCAM image where values in X-axis shows the pixel intensity whereas Y-axis represents the no of pixels.

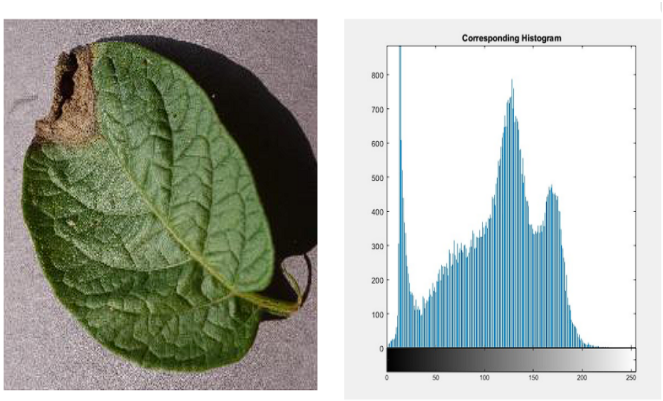

Fig. 3. Random late blight potato leaf image and its corresponding histogram

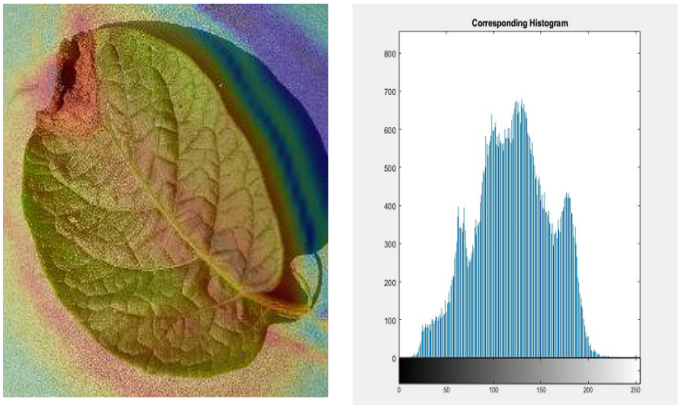

Fig. 4. GradCam image of that diseased images with its corresponding histogram

Basically, we had engaged deep learning models to identify potato leaf disease where incorporation of GRADCAM algorithm activates the salient region of the input image that is shown by the heatmap for better interpretation. Figure 3 demonstrates a sample late blight potato leaf image and its corresponding histogram are shown as well. Using GRADCAM

techniques, a GRADCAM image of heatmap had been generated from that image highlighting the salient regions which contribute to the prediction and its corresponding histogram is shown in Figure 4 From both the histograms of the input image and it's corresponding GRADCAM image of heatmap, we can see the change of number of pixels from one to another based on which the interpretation can be made.

V. Conclusion and Future Works

This research compared the classification accuracy of two established models, VGG19 and Xception, for potato disease detection. Grad-CAM was also used in the study to provide clarity on the models' interpretability for their classifications. While the models effectively identified late blight with minimal computational effort, their accuracy for classifying healthy and diseased leaves ranged from 78.47% to 83.54%. This lower accuracy is likely attributed to two factors: insufficient data and class imbalances within the dataset. To address these limitations, we have plans to integrate additional datasets into the system, potentially improving the models' performance. Furthermore, future work will focus on overcoming the current limitations and potentially incorporating automatic disease severity estimation into the system.

References

[1] Ristaino, J. B., Anderson, P. K. (2021). Persistent global change: the importance of local and historical context. Journal of Plant Diseases and Protection, 128(2), 253-260.

[2] Chakraborty, S., Al-Khuzai, A., Liu, Y., Parisi, D. (2021). Automated Detection of Potato Early and Late Blight Diseases Using Deep Learning. In 2021 IEEE International Conference on Big Data (Big Data) (pp. 1979-1987). IEEE.

[3] Sinshaw, A., Wibowo, Y., Herwanto, M. (2022). Applications of Deep Learning to Detect Potato Diseases. Journal of Physics: Conference Series, 2217(1), 012041.

[4] Islam, M. S., Razia, S., Ullah, M. A. (2017). Detection and classification of potato diseases using image segmentation and SVM. Journal of theoretical and applied information technology, 95(20), 5883-5891.

[5] Islam, M. S., Mahbub, S., Rahman, M. M. (2022). Deep learning model based approach to detect and classify potato diseases. Journal of Scientific and Industrial Research, 81(4), 346-349.

[6] Mehedi, S. H., Nisha, S. S., Alam, M. (2022). Plant disease detection using deep learning with explainable AI framework. Journal of Computer and Communications, 10(2), 23.

[7] Moin, A. (2022). Bangladeshi crops disease dataset.

[8] Simonyan, K., Zisserman, A. (2015). Deep Convolutional Neural Networks for Large-Scale Image Recognition. CoRR, abs/1409.1556.

[9] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1251-1258). IEEE.

[10] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In 2017 IEEE International Conference on Computer Vision (ICCV) (pp. 618-626). IEEE.