

Explainable ensemble deep learning for potato leaf pest and diseases identification

Nabila Husna Shabrina^{1*}, Siwi Indarti², Irmawati³, Dinar Ajeng Kristiyanti³, Niki Prastomo⁴, Tika Adillah M⁵

¹ Department of Computer Engineering, Universitas Multimedia Nusantara, Tangerang, Indonesia,
e-mail: nabila.husna@umn.ac.id

² Department of Plant Protection, Universitas Gadjah Mada, Yogyakarta, Indonesia,
e-mail: siwi.indarti@ugm.ac.id

³ Department of Information System, Universitas Multimedia Nusantara, Tangerang, Indonesia,
e-mail: irmawati@umn.ac.id ; dinar.kristiyanti@umn.ac.id

⁴ Department of Physics Engineering, Universitas Multimedia Nusantara, Tangerang, Indonesia,
e-mail: niki.prastomo@umn.ac.id

⁵ Department of Information Technology, Universitas Bina Sarana Informatika, Jakarta Pusat, Indonesia
e-mail: tika.tam@bsi.ac.id

Abstract

Potatoes are one of the most important food crops globally, yet their susceptibility to a wide range of diseases poses a serious threat to agricultural productivity. Conventional methods for identifying potato diseases are labor-intensive and time-consuming, underscoring the need for more advanced solutions. Deep learning has emerged as a promising alternative, outperforming traditional methods with its ability to automate and enhance potato disease identification. However, previous research has primarily focused on single-dataset implementations, and the interpretability of these models remains insufficiently explored. To address this, we introduce a solution that combined the strengths of ensemble deep learning and Explainable AI (XAI) for the identification of potato leaf pests and diseases. Our ensemble model integrates MobileNetV3-Large and EfficientNetV2B3 architectures and demonstrates remarkable accuracy of 97.91%. Furthermore, the incorporation of XAI techniques greatly enhances the interpretability of the model. By improving both interpretability and predictive accuracy, these results support more informed and reliable model outputs.

Keywords: ensemble deep learning, explainable artificial intelligence, image classification, potato leaf pest and disease identification

1 Introduction

Potato is the third most important food crop globally after wheat and rice, and it is a primary source of food for over a billion people worldwide. It is a crucial source of calories and nutrition, with an annual production of over 300 million metric tons ^[1]. Potatoes are also used as an industrial raw material. Maintaining potato plant health is important for ensuring food security because diseases can lead to a significant reductions in crop yield ^[2]. Plant disease experts and experienced farmers play a crucial role in monitoring crops and issuing timely warnings to mitigate potential yield losses ^[3]. However, traditional methods for disease detection in agriculture are often time-consuming, labor-intensive, and inefficient, relying heavily on manual inspection, which can be tedious and prone to human error ^[4]. Therefore, rapid and accurate identification of diseases is crucial for effectively addressing these problems.

Machine Learning (ML) and image analysis are increasingly being viewed as viable alternatives for the continuous monitoring of plant diseases across various crops ^[5,6]. Although notable successes have been reported in the literature, conventional ML methods have several limitations, including reliance on manually designed features, complex image processing steps, and a lack of robustness. The advancement of Deep Learning (DL) has provided a powerful approach to address this challenge. In particular, Convolutional Neural Networks (CNN) have demonstrated exceptional performance in image classification tasks, which has led to their use in agriculture ^[7–17]. In potato leaf disease detection, DL has demonstrated remarkable success, achieving superior accuracy and reliability compared to traditional methods. By leveraging advanced architectures, DL models can automatically extract features from leaf images, thereby enabling fast and accurate identification ^[18–22].

However, previous studies have been limited to implementation on a single dataset, which hinders the models' generalizability. Moreover, their lack of interpretability and explainability has not been adequately investigated. This limitation hinders the practical application of the existing models. To overcome this challenge, we developed a novel approach for detecting potato leaf pests and diseases. By leveraging an ensemble deep learning model combined with explainable AI (XAI) ^[23], our method not only enhances detection accuracy but also identifies the most critical features, offering valuable insights for practical agricultural use. Our objective is to aid decision-making in agriculture through an explainable deep-learning-guided method. The proposed model was trained on multiple datasets to address the limitations of prior research, which lacked generalizability due to training on a single dataset. The primary contributions of this study are as follows:

- Integration of multiple datasets to enhance the model's generalizability and robustness across diverse conditions.
- Development of an explainable deep learning framework for identifying potato leaf pests and diseases, using an ensemble of MobileNetV3 Large ^[24] and EfficientNetv2B3 ^[25].
- Implementation of explainable AI (XAI) techniques to ensure transparency and provide deeper insights into the decision-making process of the proposed model.

2 Related Work

Potato leaf disease identification has garnered significant attention recently because of its critical impact on agricultural productivity and food security. Researchers are increasingly adopting ML and DL techniques to speed up identification processes. Islam

et al. developed a potato disease recognition model that uses image processing and ML techniques to identify over 300 images using PlantVillage Dataset and achieved 95% accuracy ^[19]. Patil et al. used PlantVillage and a self-collected dataset to conduct a comparative analysis using ANN, RF, and SVM methods to identify potato disease images; RF achieved 79% accuracy, SVM achieved 84%, and ANN gained the highest accuracy of 92% ^[18]. Ji et al. used a hyperspectral imaging technique and discrete wavelet transform to recognize bruised potatoes, achieving 99.82% accuracy for damaged potatoes ^[26].

A customized and dedicated CNN was also developed specifically for potato farming. Adillah M and Kristiyanti implemented pre-trained MobileNetV2 using transfer learning and augmentation techniques on the PlantVillage Dataset. The proposed method yielded an accuracy of more than 90% ^[20]. Chen et al. developed a weakly supervised learning framework for identifying potato plant diseases. The proposed framework utilizes a foundation network and a modified version of MobileNet V2 to enhance the ability of the architecture to detect small lesions. The proposed method achieved average accuracy and specificity of 97.33% and 98.39%, respectively, on a dataset of locally sourced images ^[21]. Firasari and Cahyanti employed a PLD dataset and utilized a CNN model optimized with the RMSProp algorithm. The proposed method achieving an impressive accuracy of 97.53% ^[22].

The hybrid plant disease classification model proposed by Tabbakh et al. utilized a Transfer Learning-based model followed by a vision transformer (TLMViT). The proposed model was used to extract deep features and classify diseased plant leaves, including potato leaves. The results demonstrated that TLMViT outperformed the transfer learning-based model for PlantVillage by achieving an enhancement of 1.11% and 1.099% in validation accuracy, and 2.576% and 2.92% in validation loss, respectively ^[27]. Chen et al. introduced MobOcaNet, a novel network architecture designed for the identification of potato diseases. Building on the foundation of MobileNetV2, a lightweight network, the authors enhanced its ability to detect small crop lesions by incorporating an attention mechanism and an octave convolution block. The proposed method demonstrated superior performance, achieving an average accuracy of 97.73% in identifying various types of potato diseases, outperforming the other methods ^[28]. Javed Rashid et al proposed multi-level DL techniques by integrating YOLOv5 for leaf segmentation and a novel convolutional neural network (PDDCNN) for disease detection, achieving a high accuracy of 96.71% on the PlantVillage dataset ^[29].

To address the "black box" nature of deep learning models, XAI methods were developed to interpret model decisions. In agriculture, these methods have been applied to tasks such as plant disease detection ^[30], crop recommendation ^[31], yield estimation ^[32], and nutrient deficiencies identification ^[33]. Visualization techniques such as SHAP and Grad-CAM are particularly popular for enhancing model interpretability and transparency ^[30,32,33]. In the context of potato leaf disease identification, Bengamra et al. used saliency explanations to highlight the relevant regions of input images.

Leveraging XAI techniques empower farmers with interpretable AI, driven insights, foster trust and encourage the adoption of AI technologies in agriculture ^[34]. However, the use of XAI for potato leaf disease identification remains limited. To address this gap, the present study aims to integrate XAI techniques to improve the transparency of model predictions while simultaneously enhancing the performance of deep learning models for identifying potato leaf diseases. In addition, unlike previous studies that focused on a single dataset, this study leveraged multiple datasets to increase the model's generalizability.

3 Materials and Method

3.1 Research Workflow

Fig. 1 illustrates the workflow on the development of the proposed methods for potato leaf pest and disease datasets. The process began with the acquisition of two public datasets from the PlantVillage dataset^[35] and the Potato Leaf Dataset (PLD)^[29]. The dataset was then expanded through data augmentation, which involved the use of multiple techniques to increase its size. These methods include adjusting the brightness, flipping images vertically and horizontally, changing the zoom range, rotating the picture, and shifting it in both the length and width dimensions. Then, an ensemble model was formulated to integrate the predictions of the combined dataset, thereby enhancing the overall accuracy. Moreover, an average and concatenated learning ensemble model was formed by utilizing pre-trained transfer learning models, namely MobileNetV3 Large^[24] and EfficientNetV2B3^[25,36], to generate more accurate predictions. The resulting ensemble model was evaluated using various metrics, including accuracy, recall, precision, and F1 Score. Finally, an explanation based on the XAI algorithms was applied to interpret and explain the predicted results.

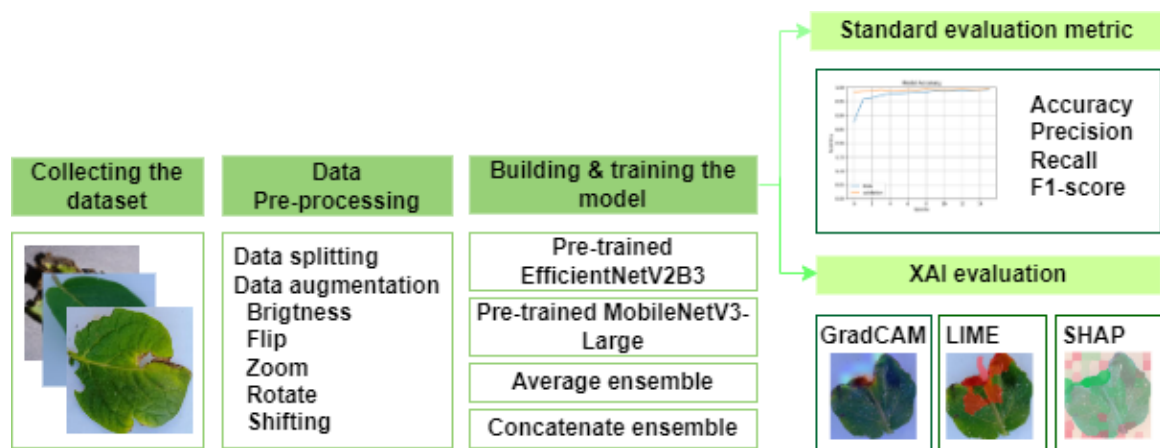


Figure 1. Schematic of the research workflow







3.2 Dataset

The datasets used in this study were sourced from the PlantVillage dataset^[35] and the Potato Leaf Dataset (PLD)^[29]. The PlantVillage dataset was developed by Penn State University (United States), and EPFL (Switzerland), and contained approximately 2,152 images for training and testing. The PLD was developed in Pakistan's Central Punjab region and comprises approximately 3,251 training and 405 testing images. In summary, the datasets comprised of 5,187 and 621 images for training and testing, respectively. The datasets were divided into three classes: early blight, late blight, and healthy. The two datasets were combined and trained using the proposed ensemble deep-learning model. The amount of data per class for training and testing is presented in Table 1. Table 2 presents a sample of potato leaf pest and disease images from both datasets.

Table 1. Distribution of potato leaf pest and diseases dataset

Classes	Training		Testing	
	PlantVillage	PLD	PlantVillage	PLD
Early Blight	900	1,303	100	162
Late Blight	900	1,132	100	141
Healthy	136	816	16	102
Total	1,936	3,251	216	405

Table 2: Sample of the potato leaf pest and diseases dataset

Category	PlantVillage dataset	PLD
Late blight		
Early blight		
Healthy		

3.3 Data Preprocessing and Bias Mitigation

Figure 2 shows the UMAP^[37] visualization of feature embeddings derived from the PlantVillage (red) and PLD (blue) datasets. The clear separation between the two clusters highlights a significant domain shift between the datasets. This domain shift can be attributed to differences in image acquisition conditions, environmental backgrounds, and disease expression across potato varieties. The PlantVillage dataset, collected in a controlled laboratory environment, forms a tight and uniform cluster, whereas the PLD dataset appears more dispersed and distinct. To address domain shifts between the PlantVillage and PLD dataset, we applied a series of data augmentation techniques, including brightness adjustment, flipping, zooming, rotation, and shifting.

The applied augmentations were intended to replicate typical conditions encountered in agricultural environments, including variations in lighting, leaf orientation, and image

capture angles. While they may not fully replicate extreme scenarios, such as severe lighting conditions or advanced stages of disease, they effectively represent a broad range of realistic variations. Additionally, these augmentation strategies were employed not only to simulate field variability but also to address class imbalance within the datasets, ensuring more balanced learning across disease categories.

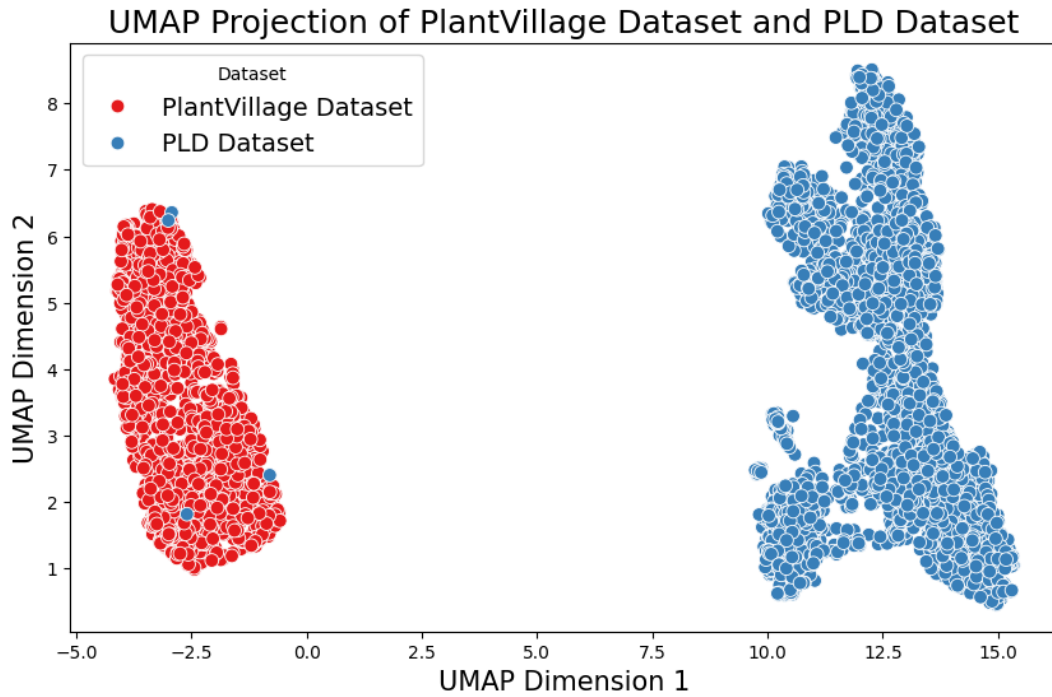


Figure 2. UMAP visualization shows domain shifts between PlantVillage and PLD dataset.

The augmentation was performed only on the training set of the dataset. This process entailed enlarging each class until it comprised 3,214 early blight classes, 3,253 late blight classes, and 3,289 healthy classes, resulting in a total of 9,756 images. This amount was split into training and validation sets at a ratio of 90:10, consisting of 8,781 images and 975 images for training and validation, respectively. The model was evaluated on 621 images from the test dataset. Below are the details regarding the implementation of data augmentation on the training dataset.

- The brightness of the images was adjusted randomly from 0.5 to 1.2.
- The images were subjected to horizontal and vertical flipping at 180 degrees.
- The degree of rotation of the images varied within 25 degrees.
- The zoom range was randomly changed from 0.6 to 0.9.
- The images were also shifted horizontally and vertically in the range 0.1.

To further enhance robustness and geographic coverage, we combined both datasets and applied transfer learning^[38] as a domain adaptation technique before integrating the models into an ensemble learning framework.

3.4 Classification System Design using Ensemble Learning

The classification system was based on two pretrained models: MobileNetV3-Large and EfficientNetV2B3. The MobileNetV3-Large and EfficientNetV2B3 models were

selected owing to their lightweight nature and outstanding performance. Tables 3 and 4 present the model architectures of MobileNetV3-Large and EfficientNetV2B3.

Table 3. MobileNetV3-Large Architecture

Operator	Input	Expansion size	Stride
Conv2d	$224^2 \times 3$	-	2
Bneck 3x3	$112^2 \times 16$	16	1
Bneck 3x3	$112^2 \times 16$	64	2
Bneck 3x3	$56^2 \times 24$	72	1
Bneck 5x5	$56^2 \times 24$	72	2
Bneck 5x5	$28^2 \times 40$	120	1
Bneck 5x5	$28^2 \times 40$	120	1
Bneck 3x3	$28^2 \times 40$	240	2
Bneck 3x3	$14^2 \times 80$	200	1
Bneck 3x3	$14^2 \times 80$	184	1
Bneck 3x3	$14^2 \times 80$	184	1
Bneck 3x3	$14^2 \times 80$	480	1
Bneck 3x3	$14^2 \times 112$	672	1
Bneck 5x5	$14^2 \times 112$	672	2
Bneck 5x5	$7^2 \times 160$	960	1
Bneck 5x5	$7^2 \times 160$	960	1
Conv 2d, 1x1	$7^2 \times 160$	-	1
Pool 7x7	$7^2 \times 960$	-	1
Conv2d 1x1, NBN	$1^2 \times 960$	-	1
Conv2d 1x1, NBN	$1^2 \times 1280$	-	1

Table 4. EfficientNetV2B3 architecture

Operator	Channel	Layer	Stride
Conv3x3	40	1	1
Fused-MBConv1, 3x3	16	1	2
Fused-MBConv4, 3x3	40	3	2
Fused-MBConv4, 3x3	56	3	2
MBConv4, 3x3, SE 0.25	112	5	1
MBConv6, 5x5, SE 0.25	136	7	2
MBConv6, 3x3, SE 0.25	232	12	1
Conv1x1 & Pooling & FC	1536	1	1

Using two pretrained models, average and concatenate ensemble techniques were constructed. The average ensemble technique was implemented by computing the average of the prediction outputs of two pretrained models that had previously been trained. This process generated new ensemble predictions that could be retrained. A visual representation of this technique is shown in Fig. 3. In the average ensemble method, the final prediction is obtained by averaging the outputs of the multiple models. Suppose we have N models, and each model i produces a prediction y_i for a given input. The ensemble prediction \hat{y} is calculated as in Eq. (1).

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (1)$$

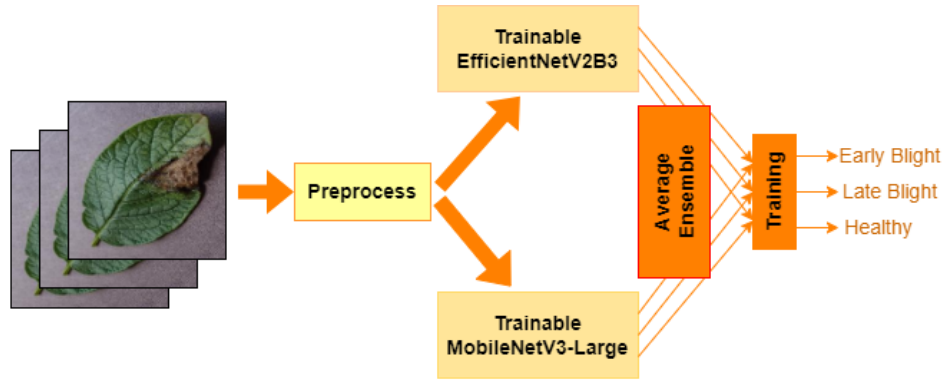


Figure 3. The proposed average ensemble deep learning

The concatenation ensemble involves combining the features of two base pretrained models into a single, larger feature, which is then subsequently employed in the classification layer. The scheme employed in the concatenation ensemble is illustrated in Fig. 4. The concatenate ensemble method combines the feature representations learned by different models prior to making a final prediction. For two models, each producing a feature vector f_1 and f_2 , respectively, the concatenated feature vector f_{concat} is formed as in Eq. (2). The $[\cdot]$ denotes the concatenation operation. The combined feature vector f_{concat} is fed into a subsequent classifier (e.g., a fully connected neural network) to produce the final prediction. The proposed method leverages the strengths of each model's feature extraction capabilities to capture more comprehensive information from the input data.

$$f_{concat} = [f_1; f_2] \quad (2)$$

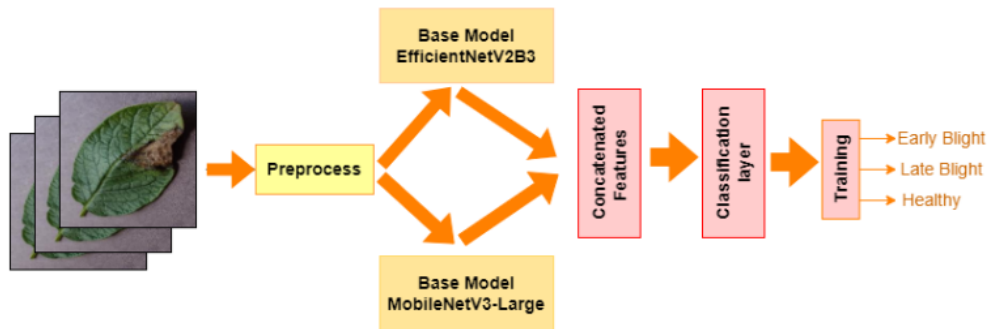


Figure 4. The proposed concatenate ensemble deep learning

The average and concatenate ensemble methods were chosen over alternatives like boosting or bagging because they are well-suited for combining pretrained deep learning models, particularly in tasks like image classification^[39]. Boosting methods, typically rely on iterative training multiple weak learners to correct errors in previous iterations, which is not as effective for complex deep learning models where pre-trained architectures already capture a high level of feature representation^[40]. Similarly, bagging approaches focus on training multiple independent models on different subsets of data, which may not effectively leverage the strengths of pre-trained models effectively in this context.

The average and concatenate ensembles allow us to leverage the unique strengths of both EfficientNetV2B3 and MobileNetV3-Large without requiring complex iterative

processes. By combining these two models, the ensemble method captures both the robust feature extraction capabilities of EfficientNetV2B3 and the computational efficiency of MobileNetV3-Large. The average ensemble integrates the prediction probabilities from both models, reducing the individual model biases and providing balanced and robust output. In addition, the concatenate ensemble fuses the feature representations of both models, thereby creating a richer feature space that allows the final classifier to learn patterns in the data.

3.5 Hyperparameter and Evaluation Metrics

The use of hyperparameters is important for attaining desired performance results. Table 5 presents the details of the employed hyperparameters and their respective functions. The Adam optimizer was selected because it combines the benefits of both the Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp)^[41], thereby offering efficient performance across a wide range of deep learning tasks. A low learning rate of 0.0001 was selected to ensure stable and gradual convergence during training. Categorical crossentropy was used as the loss function because the task involved multi-class classification of potato leaf diseases. A batch size of 64 was selected to balance computational efficiency and gradient stability. Finally, the model was trained for 50 epochs to provide sufficient data exposure, which enable effective learning while minimizing the risk of overfitting.

Table 5. Hyperparameter settings

Hyperparameters	Value
Optimizer	Adam
Learning Rate	0.0001
Losses	Categorical Crossentropy
Batch size	64
Epoch	50

The evaluation metrics of the trained models play a crucial role in their analysis, serving to assess performance, compare models, optimize the model, and draw conclusions. In this study, several standard evaluation metrics were employed, namely, accuracy, recall, precision, and F1 score. The formula for each metric is given by Eqs. (3)-(6).

$$\text{Test Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (3)$$

$$\text{Precision} = \frac{TP}{TP+FP}, \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN}, \quad (5)$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (6)$$

where TP is a True Positive; FP is a False Positive; TN is a True Negative; FN is a False Negative. The utilization of the learning curve was also incorporated into the evaluation of the model's performance in addition to the standard evaluation metric.

3.6 Explainable-driven method

Several XAI techniques were applied to explain the predicted results. XAI encompasses a range of techniques and approaches aimed at increasing the transparency and understandability of the decisions and outputs of model systems. In this study, LIME^[42], SHAP^[43], and GradCAM^[44], three commonly used XAI methods, were employed. The GradCAM technique was used to identify the vital regions of the potato leaf images necessary for classification by utilizing the spatial information maintained by the convolutional layers. To assess the usefulness of the proposed visual explanation methods, a thorough examination of individual potato leaf samples from each category was carried out, including a visual inspection of the heatmaps generated by the methods.

Shapley Additive Explanation (SHAP) is similar to the method used in game theory to enhance the interpretability of specific predictions by determining the significance values for each input feature. This provided a description of the predicted outcome based on the contribution of each feature. This technique has also been applied to surrogate models^[45]. It is easier to calculate and provides more natural interpretations than other methods^[46]. SHAP generates both local and global explanations, making it a reliable option for any data, and it is not limited to model-agnostic situations.

Local interpretable model-agnostic explanation (LIME) is a model-agnostic tool generates local explanations of a model's predictions by identifying the most relevant features necessary for the prediction. Unlike SHAP, LIME does not rely on game theory, instead, it uses a direct approach by varying the input data of the model to observe changes in the prediction. The explanations provided by LIME are based on individual instances rather than the entire dataset, and they train the model locally to provide explanations for each prediction. In addition, LIME employs hierarchical clustering to select the most relevant cluster of instances for explanation^[47,48].

3.7 Implementation

Using Keras and the TensorFlow Library in Python, the ensemble deep learning model and XAI were implemented in Google Colab with the specifications of the Tesla K80 accelerator, CPU Xeon Processor at 2.2 GHz based on availability, and 12 GB RAM. The training process also incorporated custom callbacks, including the implementation of checkpoints and early stopping functions, to ensure efficient and effective model training. A learning rate scheduler was also employed to optimize the training process and mitigate overfitting and instability.

4 Results and Discussion

4.1 Evaluation of Ensemble Deep Learning Model Performance

Figs. 5 and 6 show the learning curve results from the pretrained MobileNetV3-Large and EfficientNetV2B3 models, respectively. The learning curves for the proposed average ensemble and concatenated ensemble are shown in Figs. 7 and 8, respectively.

Figs. 4 and 5 demonstrate that both the pretrained deep learning models yielded excellent results, with validation and training accuracy exceeding 95%. In addition, the convergence between the two accuracy was noteworthy. When comparing the performance of EfficientNetV2B3 and MobileNetV3Large, EfficientNetV2B3 outperforms MobileNetV3Large. In particular, EfficientNetV2B3 achieved a maximum validation

accuracy of 99.08%, which is higher than the training accuracy of 97.85%. In contrast, MobileNetV3Large's maximum validation accuracy of 98.97% is lower than its training accuracy of 99.32%.

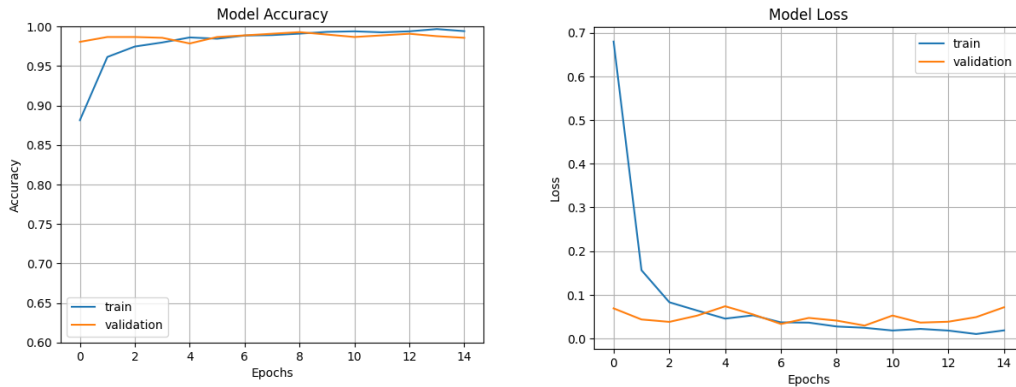


Figure 5. Model accuracy and loss of MobileNetV3-Large

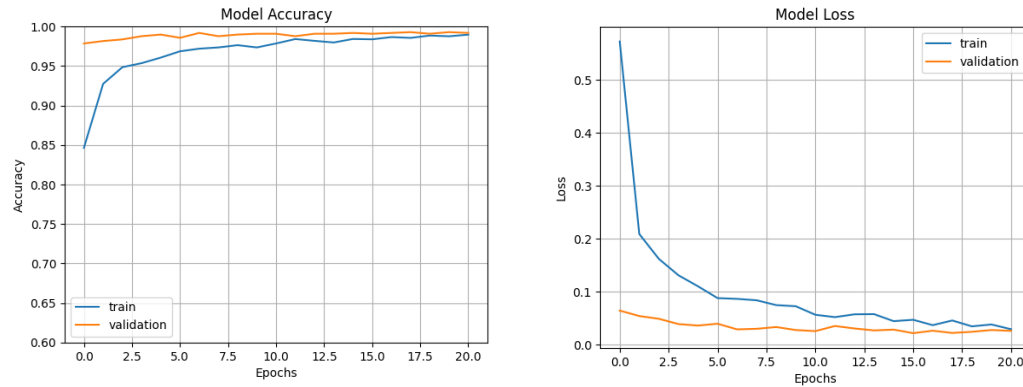


Figure 6. Model accuracy and loss of EfficientNetV2B3

As shown in Fig 7. And Fig. 8, the concatenate ensemble resulted in training, validation, and testing accuracy values of 99.25%, 99.49%, and 97.91%, respectively. The average ensemble deep learning achieved training, validation, and testing accuracies of 97.55 %, 98.55 %, and 97.91 %, respectively. The use of the average ensemble deep learning model did not satisfy expectations because the average performance of the ensemble model was subpar. The resulting graphs do not converge. This implies that the model faces an underfitting problem. The curve suggests that the model has potential for further development and enhancement; however, the training process was discontinued before its full potential. The addition of network complexity may overcome this problem. However, the concatenated ensemble model outperformed both EfficientNetV2B3 and MobileNetV3Large. The validation accuracy was 99.18% and the training accuracy was 98.69%. In addition, the validation and training accuracy were converged more quickly using the concatenated ensemble model. This indicates that the proposed concatenation ensemble deep learning method demonstrates potential and may prove to be highly effective for the identification of potato leaf pests and diseases.

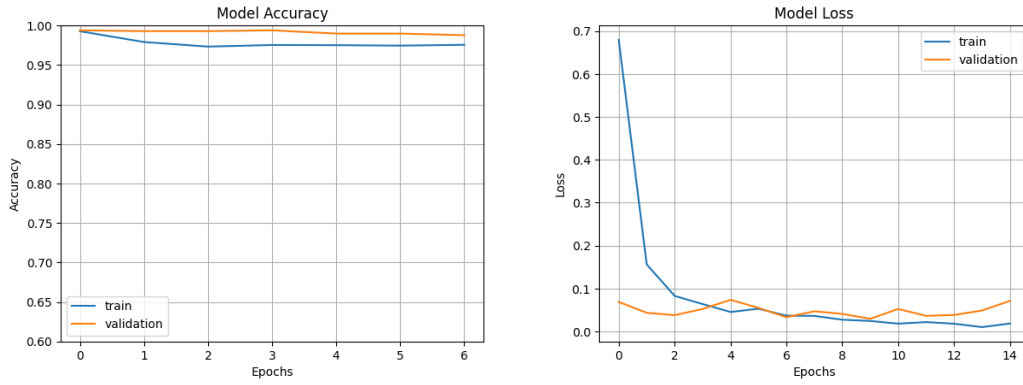


Figure 7. Model accuracy and loss of the proposed average ensemble

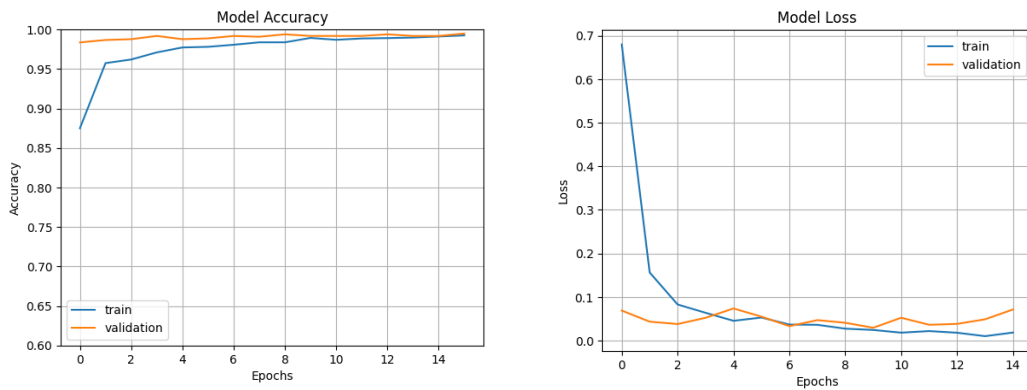


Figure 8. Model accuracy and loss of the proposed concatenate ensemble

The performance of the proposed method was further verified by calculating the precision, recall, and F1-score, as listed in Table 6. In terms of accuracy, both average and concatenated ensemble deep learning demonstrated superior performance compared to the pretrained EfficientNetV2B3 and MobileNetV3-Large, with values of 97.91%. The concatenated ensemble also achieved an impressive average recall score of 98.1%, which significantly surpassed that of the other models. The precision and F1-score of the average ensemble learning achieved the highest results 97.37% and 97.6%, respectively. In conclusion, the results of the experiment revealed that despite its lackluster performance in the learning curve, the average ensemble deep learning demonstrated superior performance in terms of testing accuracy, precision, and F1-score. Furthermore, the proposed concatenation approach demonstrated promising results, achieving noteworthy performance in both the learning curve and all metric evaluations, compared to the pretrained EfficientNetV2B3 and MobileNetV3-Large.

Table 6. Model performance result (in %)

Model	Precision	Recall	F1-score	Accuracy
EfficientNetV2B3	96.93	98	97.42	97.75
MobileNetV3-Large	96.78	97.69	97.21	97.58
Average Ensemble	97.39	97.82	97.6	97.91
Concatenate Ensemble	97.09	98.1	97.56	97.91

The results also show that both the average and concatenate ensemble methods achieve the same accuracy of 97.91%, with minor variations in precision, recall, and F1-score. This comparable performance can be attributed to the complementary strengths of the base models (EfficientNetV2B3 and MobileNetV3-Large) which are effectively integrated in both ensemble strategies. The average ensemble method averages the prediction probabilities of the base models. This ensures that each model contributes equally, effectively smoothing out individual biases and improving the overall predictions robustness. On the other hand, the concatenate ensemble method combines the feature representations from both models before making predictions, which allow the model to learn richer and more diverse features from the input data.

However, as shown in Table 6, the average ensemble produced more stable results across metrics. This can be explained by the concatenate ensemble's increased feature dimensionality, which may introduce redundancy or noise, increasing the risk of overfitting, particularly when the training data are limited or imbalanced. In contrast, the simpler structure of the average ensemble reduces variance while maintaining strong performance, which results in balanced performance across metrics.

The use of both ensemble methods highlights the robustness and versatility of combining multiple models, which results in improved performance over that of individual models. The shared accuracy between the two approaches demonstrates the effectiveness of the ensemble strategy in maximizing the predictive performance for potato leaf disease detection.

4.2 Effectiveness of the Augmentation Techniques

To validate the effectiveness of data augmentation, we conducted a comparative analysis of model performance with and without augmentation, as presented in Table 7.

Table 7. Model performance before and after augmentation (in %)

Model	Precision	Recall	F1-score	Accuracy
Average Ensemble (Before Augmentation)	96.04	96.46	96.22	96.51
Average Ensemble (After augmentation)	97.39	97.82	97.60	97.91
Concatenate Ensemble (Before Augmentation)	96.75	96.67	96.69	96.67
Concatenate Ensemble (After Augmentation)	97.09	98.1	97.56	97.91

The results presented in Table 7 demonstrate that data augmentation led to consistent improvements across all evaluation metrics for both ensemble strategies. For the Average Ensemble, precision increased from 96.04% to 97.39%, recall from 96.46% to 97.82%, and F1-score from 96.22% to 97.60%, with overall accuracy improving from 96.51% to 97.91%. Similarly, the Concatenate Ensemble showed enhancement, with F1-score rising from 96.69% to 97.56%, and accuracy increasing from 96.67% to 97.91%. These improvements indicate that augmentation not only enhances generalization but also helps the model better capture relevant features across diverse conditions. Overall, the results validate that data augmentation plays a significant role in boosting classification performance and mitigating dataset limitations.

4.3 Computational Cost for Ensemble Deep Learning

To assess the computational costs associated with each ensemble method, we performed a detailed analysis of inference time and memory consumption, as presented in Table 8. The Concatenate Ensemble achieved a slightly faster average inference time of 1.3287 seconds compared to 1.3783 seconds for the Average Ensemble. However, this performance gain came at the cost of higher memory usage, with the Concatenate Ensemble consuming 7.0 GB, whereas the Average Ensemble used only 5.4 GB. These results suggest a trade-off between speed and memory efficiency, where the Concatenate Ensemble is more time-efficient, while the Average Ensemble is more memory-efficient.

Table 8. Computational cost for ensemble deep learning


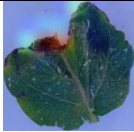



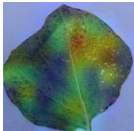
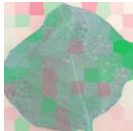


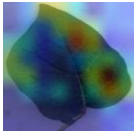
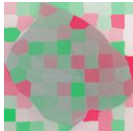

Model	Inference time (seconds)	Memory Usage (GB)
Average Ensemble	1.3783	5.4
Concatenate Ensemble	1.3287	7.0

Considering both inference time and memory usage, the choice between ensemble methods depends on the deployment context. If computational resources are limited, such as in mobile or edge devices, the Average Ensemble is more suitable due to its lower memory consumption. However, if faster inference is prioritized and sufficient memory is available, the Concatenate Ensemble offers better time efficiency. Overall, both methods perform comparably in accuracy, the decision for deployment should be based on the target application's hardware constraints and performance requirements.

4.4 Model Explanation with Ensemble Deep Learning

The predicted probabilities for each class of sample images and the results of the XAI techniques are presented in Table 9. As shown in the table, the three methods employed different analysis method. The Grad-CAM heatmap shows the significance of each spatial location in the input image for target class prediction. Warmer regions indicate higher importance, whereas cooler regions indicate lower importance. The SHAP visualization represents positive features, that influence the model's prediction and increase its output, with green color, whereas negative features are represented in red. The LIME visualization utilizes red color to guide the model to yield its predicted results. As shown in Table 9, the prediction outcomes of the proposed ensemble models mostly conformed to the expected features. Compared to the two XAI models, the Grad-CAM had poorer evaluation results. The evaluation provided by Grad-CAM still provides a background that should not be the main feature of the late blight disease class. LIME improves the evaluation of Grad-CAM by reducing background noise, while SHAP provides more detailed and precise evaluation results.

Table 9. XAI results

Actual class	Prediction result	GradCAM	SHAP	LIME
 Late blight	Late Blight (95.29%)			
 Early blight	Early Blight (99.9%)			
 Healthy	Healthy (100%)			

To complement the qualitative results, we provide a quantitative comparison of the three XAI methods, as presented in Table 10. All methods demonstrated high fidelity, with LIME achieving the highest score (1.0000), followed closely by GradCAM (0.9958) and SHAP (0.9747). These results indicate that all three techniques can represent the model's predictions effectively. However, a significant variation was observed in inference time. SHAP was the fastest, requiring only 5 seconds, whereas GradCAM and LIME took 210 seconds and 268 seconds, respectively. This result highlights a trade-off between interpretability fidelity and computational efficiency, where SHAP offers rapid explanations with slightly lower fidelity, while LIME provides the most reliable interpretation at the expense of processing time.

Table 10. Quantitative comparison of the XAI method

XAI Method	Fidelity score	Inference time (seconds)
GradCAM	0.9958	210
SHAP	0.9747	5
LIME	1.0000	268

This study effectively integrates XAI techniques like Grad-CAM, SHAP, and LIME, to improve the interpretability of model predictions. While the XAI methods evaluated in this study demonstrated high fidelity, there are significant differences in computational efficiency. SHAP and LIME took more than 200 seconds for inference, presenting a computational bottleneck that may hinder scalability and real-time deployment in resource-constrained settings. To address this issue, future work could focus on exploring other XAI techniques, such as MASHAP^[49], or FastSHAP^[50]. Another potential approach could involve applying those techniques selectively to smaller subsets of the dataset or reducing the dimensionality of input features, which can help to minimize computational demands without sacrificing interpretability.

4.5 Comparison with Other Studies

To evaluate the effectiveness of our proposed method, we compared our results with those of similar studies. The comparison in Table 11 demonstrates the performance of various methods for potato leaf disease detection, with our proposed approach achieving the highest accuracy of 97.91%. Our proposed method stands out not only for its superior accuracy but also for its ability to generalize effectively across multiple datasets, including PlantVillage and PLD. This highlights the advantage of leveraging an ensemble of MobileNetV3 and EfficientNetV2B3, which combines the strengths of both architectures to enhance feature extraction and classification performance.

Table 11. Comparison with other studies

Author	Method	Dataset	Accuracy
Islam et al. ^[19]	Segmentation & SVM	PlantVillage	95%
Patil et al. ^[18]	ANN	PlantVillage & Self Collected dataset	92%
Firasari and Cahyanti ^[22]	CNN with RMS Prop Optimizer	PLD	97.53%
Rashid et al. ^[29]	Multilevel DL	PlantVillage	96.71
Ours	Ensemble MobileNetV3 and EfficientNetV2B3	PlantVillage and PLD	97.91%

However, we acknowledge that there are limitations in the comparability of these results due to inconsistent dataset splits and evaluation protocols used across the referenced studies. Islam et al. used a 60%-40% train-test split with 300 images from PlantVillage, while Rashid et al. employed an 80%-10%-10% train-validation-test division. Patil et al. utilized 892 images from mixed sources, including PlantVillage, but did not mention their data split strategy. Similarly, Firasari and Cahyanti divided the PLD dataset into three subsets, though the exact proportions were not reported. Furthermore, the lack of access to original code prevented us from retraining these methods under uniform experimental conditions. Therefore, our comparison is based solely on the reported results.

4.6 Deployment Considerations

The findings of this study have the potential to be integrated into a user-friendly application aimed at assisting plant disease experts and farmers in identifying potato leaf diseases. This would involve deploying the trained model in a lightweight and platform-compatible format, such as TensorFlow Lite or ONNX, to support both mobile and web-based environments. However, despite the improved classification performance demonstrated by the ensemble models, their deployment on resource-constrained platforms presents significant challenges. Ensemble architecture generally leads to increased model size, longer inference times, and higher memory usage, which can be difficult for devices with limited computational resources.

To address these concerns, future work may focus on optimization strategies, such as model compression, quantization, and pruning, to reduce latency and resource usage. In

addition, selecting more lightweight base models or applying knowledge distillation to transfer ensemble knowledge into a smaller single model can make real-time deployment more feasible while preserving model accuracy. By deploying the model through a user-friendly application, users would be able to upload images directly via the app interface. The system would preprocess the input and run it through the trained model to generate real-time predictions. Additionally, the app could visually highlight the most relevant leaf regions contributing to the prediction, providing users with intuitive and transparent insights into the model's decision-making process.

4.7 Challenges and Future Opportunities

Despite the overall strong performance of the proposed model, we recognize the possibility of biases arising from underrepresented disease classes and limited geographic coverage within the datasets. Such imbalances can affect the model's ability to generalize across all disease types and farming regions. To address this, we implemented class balancing techniques, including applied data augmentation with oversampling to simulate a wider range of real-world conditions. Additionally, we combined both the PlantVillage and PLD datasets to evaluate the model's robustness across diverse domains. Moving forward, we aim to incorporate more regionally diverse data and conduct field-level validation to further ensure fairness and generalizability.

While this study focused specifically on potato leaf disease identification, the proposed model has the potential to be generalized and adapted to detect diseases in other crops. The methods used, including deep learning and explainable AI techniques, are not limited to potato leaf datasets and can be applied to similar image-based plant disease datasets. The model can be further developed by incorporating additional training data representing a diverse range of crops and their associated diseases.

5 Conclusion

We have presented a framework for generating explanations along with an ensemble deep learning model comprising MobileNetV3-Large and EfficientNetV2B3, which achieved a remarkable testing accuracy of up to 97.91% for potato leaf pest and disease identification. The ensemble model was trained and evaluated using a detailed explanation generated by the applications of GradCAM, SHAP, and LIME. The results indicate that the explanations generated were able to identify the specific regions responsible for the classification of potato leaf pests and diseases. The proposed model demonstrates the potential of XAI and ensemble deep learning for generating admissible explanations of outcomes with high classification accuracy. The findings of this study can be incorporated into a user-friendly platform designed to assist plant disease experts and experienced farmers in accurately identifying potato leaf pests and diseases. The platform will deliver interpretable, AI-powered insights, empowering users to make well-informed decisions on potato disease management and pest control. Future research will also focus on developing models using a more diverse dataset that includes a broader range of images and classes to enhance the robustness of the results.

Acknowledgements

This study was financially supported by the Ministry of Research, Technology, and Higher Education of Indonesia in Fundamental Research Schemes, Grant number

1423/LL3/AL.04/2023. The authors would also like to thank Universitas Multimedia Nusantara and Universitas Gadjah Mada for their support during this study

Data Availability

The code used in this study are publicly available at our GitHub repository: <https://github.com/nhs-research/potato-classification>. The datasets used (PlantVillage and PLD) are publicly accessible as cited in References ^[29] and ^[35].

References

- [1] Chauhan, A., Islam, F., Imran, A., Ikram, A., Zahoor, T., Khurshid, S., & Shah, M. A. (2023). A review on waste valorization, biotechnological utilization, and management of potato. *Food Science & Nutrition*, 11(10), 5773–5785. <https://doi.org/10.1002/fsn3.3546>
- [2] Fang, Y., & Ramasamy, R. (2015). Current and Prospective Methods for Plant Disease Detection. *Biosensors*, 5(3), 537–561. <https://doi.org/10.3390/bios5030537>
- [3] Venbrux, M., Crauwels, S., & Rediers, H. (2023). Current and emerging trends in techniques for plant pathogen detection. *Frontiers in Plant Science*, 14, 1120968. <https://doi.org/10.3389/fpls.2023.1120968>
- [4] Ahmad, A., Saraswat, D., & El Gamal, A. (2023). A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. *Smart Agricultural Technology*, 3, 100083. <https://doi.org/10.1016/j.atech.2022.100083>
- [5] Azlah, M. A. F., Chua, L. S., Rahmad, F. R., Abdullah, F. I., & Wan Alwi, S. R. (2019). Review on Techniques for Plant Leaf Classification and Recognition. *Computers*, 8(4), 77. <https://doi.org/10.3390/computers8040077>
- [6] Ramesh, S., Hebbar, R., M., N., R., P., N., P. B., N., S., & P.V., V. (2018). Plant Disease Detection Using Machine Learning. *2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, 41–45. <https://doi.org/10.1109/ICDI3C.2018.00017>
- [7] Kusrini, K., Suputa, S., Setyanto, A., Agastya, I. M. A., Priantoro, H., Chandramouli, K., & Izquierdo, E. (2020). Data augmentation for automated pest classification in Mango farms. *Computers and Electronics in Agriculture*, 179, 105842. <https://doi.org/10.1016/j.compag.2020.105842>
- [8] Wang, A. R., & Shabrina, N. H. (2023). A deep learning-based mobile app system for visual identification of tomato plant disease. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(6), 6992–7004. <https://doi.org/10.11591/ijece.v13i6.pp6992-7004>
- [9] Shabrina, N. H., Lika, R. A., & Indarti, S. (2023). Deep learning models for automatic identification of plant-parasitic nematode. *Artificial Intelligence in Agriculture*, 7, 1–12. <https://doi.org/10.1016/j.aiia.2022.12.002>
- [10] Shabrina, N. H., Indarti, S., Lika, R. A., & Maharani, R. (2023). A comparative analysis of convolutional neural networks approaches for phytoparasitic nematode identification. *Communications in Mathematical Biology and Neuroscience*. <https://doi.org/10.28919/cmbn/7993>
- [11] Lika, R. A., Shabrina, N. H., Indarti, S., & Maharani, R. (2023). Transfer Learning using Hybrid Convolution and Attention Model for Nematode Identification in Soil

- Ecology. *Revue d'Intelligence Artificielle*, 37(4), 945–953. <https://doi.org/10.18280/ria.370415>
- [12] Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-Time Detection of Apple Leaf Diseases Using Deep Learning Approach Based on Improved Convolutional Neural Networks. *IEEE Access*, 7, 59069–59080. <https://doi.org/10.1109/ACCESS.2019.2914929>
- [13] Alfarisy, A. A., Chen, Q., & Guo, M. (2018). Deep learning based classification for paddy pests & diseases recognition. *Proceedings of 2018 International Conference on Mathematics and Artificial Intelligence*, 21–25. <https://doi.org/10.1145/3208788.3208795>
- [14] Zhang, X., Qiao, Y., Meng, F., Fan, C., & Zhang, M. (2018). Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks. *IEEE Access*, 6, 30370–30377. <https://doi.org/10.1109/ACCESS.2018.2844405>
- [15] Lin, Z., Mu, S., Huang, F., Mateen, K. A., Wang, M., Gao, W., & Jia, J. (2019). A Unified Matrix-Based Convolutional Neural Network for Fine-Grained Image Classification of Wheat Leaf Diseases. *IEEE Access*, 7, 11570–11590. <https://doi.org/10.1109/ACCESS.2019.2891739>
- [16] Shabrina, N. H., & Brian, A. (2023). A Comparative Analysis of Pre-trained Deep Neural Networks for Mango Leaves Pests and Diseases Identification. *ICIC Express Letters, Part B : Applications*, 14(11), 1207–1215. <https://doi.org/10.24507/icicelb.14.11.1207>
- [17] Angeline, N., Husna Shabrina, N., & Indarti, S. (2023). Faster region-based convolutional neural network for plant-parasitic and non-parasitic nematode detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 30(1), 316. <https://doi.org/10.11591/ijeecs.v30.i1.pp316-324>
- [18] Patil, P., Yaligar, N., & Meena, S. M. (2017). Comparison of Performance of Classifiers—SVM, RF and ANN in Potato Blight Disease Detection Using Leaf Images. *2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 1–5. <https://doi.org/10.1109/ICCIC.2017.8524301>
- [19] Islam, M., Anh Dinh, Wahid, K., & Bhowmik, P. (2017). Detection of potato diseases using image segmentation and multiclass support vector machine. *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 1–4. <https://doi.org/10.1109/CCECE.2017.7946594>
- [20] Adillah M, T., & Kristiyanti, D. A. (2023). Implementation of Transfer Learning Mobilenetv2 Architecture for Identification of Potato Leaf Disease. *Journal of Theoretical and Applied Information Technology*, 101(16), 6273–6285.
- [21] Chen, J., Deng, X., Wen, Y., Chen, W., Zeb, A., & Zhang, D. (2023). Weakly-supervised learning method for the recognition of potato leaf diseases. *Artificial Intelligence Review*, 56(8), 7985–8002. <https://doi.org/10.1007/s10462-022-10374-3>
- [22] Firasari, E., & Cahyanti, F. L. D. (2023). CLASSIFICATION OF POTATO LEAF DISEASES USING CONVOLUTIONAL NEURAL NETWORK. *Jurnal Techno Nusa Mandiri*, 20(2), 89–94. <https://doi.org/10.33480/techno.v20i2.4655>
- [23] Ryo, M. (2022). Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. *Artificial Intelligence in Agriculture*, 6, 257–265. <https://doi.org/10.1016/j.aiia.2022.11.003>
- [24] Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. (2019). Searching for MobileNetV3. In: *Proceeding of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>

- [25] Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training. *In: Proceedings of the 38 Th International Conference on Machine Learning*, 139, 10096–10106. <https://doi.org/10.48550/arXiv.2104.00298>
- [26] Ji, Y., Sun, L., Li, Y., & Ye, D. (2019). Detection of bruised potatoes using hyperspectral imaging technique based on discrete wavelet transform. *Infrared Physics & Technology*, 103, 103054. <https://doi.org/10.1016/j.infrared.2019.103054>
- [27] Tabbakh, A., & Barpanda, S. S. (2023). A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer “TLMViT” for Plant Disease Classification. *IEEE Access*, 11, 45377–45392. <https://doi.org/10.1109/ACCESS.2023.3273317>
- [28] Chen, W., Chen, J., Zeb, A., Yang, S., & Zhang, D. (2022). Mobile convolution neural network for the recognition of potato leaf disease images. *Multimedia Tools and Applications*, 81(15), 20797–20816. <https://doi.org/10.1007/s11042-022-12620-w>
- [29] Rashid, J., Khan, I., Ali, G., Almotiri, S. H., AlGhamdi, M. A., & Masood, K. (2021). Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. *Electronics*, 10(17), 2064. <https://doi.org/10.3390/electronics10172064>
- [30] Patil, P., Pamali, S. K., Devagiri, S. B., Sushma, A. S., & Mirje, J. (2024). Plant Leaf disease detection using XAI. *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, 1–6. <https://doi.org/10.1109/AIIoT58432.2024.10574617>
- [31] Amin, M. N., Memon, S., Ali, A., Khan, H., Joshi, R., Afzal Rana, M. T., & ALsaawy, Y. (2024). Enhanced Deep Learning Based Precision Agriculture: A Decision Support System for Enhancing Crop Recommendation Accuracy Using Convolutional Neural Networks (CNN). *JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES*, 19(9). <https://doi.org/10.26782/jmcms.2024.09.00014>
- [32] Venugopal, A., Farnaghi, M., & Zurita-Milla, R. (2024). Comparative Evaluation of XAI Methods for Transparent Crop Yield Estimation Using CNN. *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, 7478–7482. <https://doi.org/10.1109/IGARSS53475.2024.10641426>
- [33] Mkhathshwa, J., Kavvu, T., & Daramola, O. (2024). Analysing the Performance and Interpretability of CNN-Based Architectures for Plant Nutrient Deficiency Identification. *Computation*, 12(6), 113. <https://doi.org/10.3390/computation12060113>
- [34] Konda, O., Sharief Mohammad, R. A., Mishra, S., Rajeev, N., & Verma, A. (2024). Harvesting Insights: Leveraging Explainable AI to Optimize Farming Practices. *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 1–8. <https://doi.org/10.1109/ACROSET62108.2024.10743799>
- [35] Hughes, D. P., & Salathe, M. (2016). *An open access repository of images on plant health to enable the development of mobile disease diagnostics* (No. arXiv:1511.08060). arXiv. <https://doi.org/10.48550/arXiv.1511.08060>
- [36] Tan, M., & Le, Q. V. (2019). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. <https://doi.org/10.48550/ARXIV.1905.11946>
- [37] McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*, 3(29), 861. <https://doi.org/10.21105/joss.00861>

- [38] Hosna, A., Merry, E., Gyalmo, J., Alom, Z., Aung, Z., & Azim, M. A. (2022). Transfer learning: A friendly introduction. *Journal of Big Data*, 9(1), 102. <https://doi.org/10.1186/s40537-022-00652-w>
- [39] Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151. <https://doi.org/10.1016/j.engappai.2022.105151>
- [40] Khan, A. A., Chaudhari, O., & Chandra, R. (2024). A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation. *Expert Systems with Applications*, 244, 122778. <https://doi.org/10.1016/j.eswa.2023.122778>
- [41] Kingma, D. P., & Ba, J. (2017). *Adam: A Method for Stochastic Optimization* (No. arXiv:1412.6980). arXiv. <http://arxiv.org/abs/1412.6980>
- [42] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier (No. arXiv:1602.04938). arXiv. <http://arxiv.org/abs/1602.04938>
- [43] Rodríguez-Pérez, R., & Bajorath, J. (2020). Interpretation of machine learning models using shapley values: Application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34(10), 1013–1026. <https://doi.org/10.1007/s10822-020-00314-0>
- [44] 44. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *International Journal of Computer Vision*, 128(2), 336–359. <https://doi.org/10.1007/s11263-019-01228-7>
- [45] Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. *Journal of Artificial Intelligence Research*, 70, 245–317. <https://doi.org/10.1613/jair.1.12228>
- [46] Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1), 18. <https://doi.org/10.3390/e23010018>
- [47] Loh, H. W., Ooi, C. P., Seoni, S., Barua, P. D., Molinari, F., & Acharya, U. R. (2022). Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, 226, 107161. <https://doi.org/10.1016/j.cmpb.2022.107161>
- [48] Sahakyan, M., Aung, Z., & Rahwan, T. (2021). Explainable Artificial Intelligence for Tabular Data: A Survey. *IEEE Access*, 9, 135392–135422. <https://doi.org/10.1109/ACCESS.2021.3116481>
- [49] Messalas, A., Aridas, C., & Kanellopoulos, Y. (2020). Evaluating MASHAP as a faster alternative to LIME for model-agnostic machine learning interpretability. *2020 IEEE International Conference on Big Data (Big Data)*, 5777–5779. <https://doi.org/10.1109/BigData50022.2020.9378034>
- [50] Jethani, N., Sudarshan, M., Covert, I., Lee, S.-I., & Ranganath, R. (2022). *FastSHAP: Real-Time Shapley Value Estimation*. <https://doi.org/10.48550/arXiv.2107.07436>

Notes on contributors

Nabila Husna Shabrina is an associate professor in the Department of Computer Engineering at Universitas Multimedia Nusantara. Her research interests include image processing and computer vision, with a focus on biological and medical applications.



Siwi Indarti is a professor in the Department of Plant Protection at Universitas Gadjah Mada. Her research interests include plant nematology and crop protection.



Irmawati is an assistant professor in the Department of Information System at Universitas Multimedia Nusantara. Her research interests include image processing, computer vision, machine learning and deep learning



Dinar Ajeng Kristiyanti is an associate professor in the Department of Information System at Universitas Multimedia Nusantara. Her research interests include sentiment analysis, text mining, feature selection, optimization, data science and machine learning



Niki Prastomo is an assistant professor in the Department of Physics Engineering at Universitas Multimedia Nusantara. His research interest includes functional materials.



Tika Adillah M is a lecturer in the Department of Information Technology at Universitas Bina Sarana Informatika. Her research interests include machine learning, deep learning, image processing, and text mining.