

Revolutionizing Crop Leaf Disease Detection: A Novel Ensemble Learning Framework Using Customized EfficientNets

Nahrin Jannat^{*,1,a}, S. M. Mahedy Hasan^{1,b}, and Minhaz F. Zibran²

¹ Rajshahi University of Engineering & Technology, Rajshahi, Bangladesh
^anahrinnipun17@gmail.com ^bmahedy@cs.e.ruet.ac.bd

² Idaho State University, ID, USA
²minhazzibran@isu.edu

Abstract. The presence of crop leaf diseases poses a persistent and significant threat to agricultural productivity and food security, especially in Bangladesh, where agriculture plays a pivotal role in the economy. Developing efficient methodologies for timely crop leaf disease detection and management becomes paramount. Nonetheless, our study addressed the hurdles in detecting crop leaf diseases, with a special emphasis on two datasets that included corn, and potato rather than relying on a single dataset. While existing studies often rely on straightforward transfer learning (TL) techniques, our research aimed to enhance the performance of TL by systematically incorporating various EfficientNets versions and customizing their architecture with different layers. In addition, our research made a notable contribution by introducing a novel model selection method for ensemble learning that extended beyond traditional accuracy metrics to address misclassifications and class-specific gaps. We devised a customized approach incorporating misclassification counts and Hamming Loss, redefining the model selection process. Furthermore, we identified the most suitable EfficientNets models for each dataset and leveraged the **Gradient Class Activation Map (Grad-CAM)** for decision visualization of the model. Consequently, our research effectively addressed agricultural challenges and paved the path toward more robust and precise crop leaf disease detection.

Keywords: Crop Leaf · Disease · EfficientNet · Misclassification · Hamming Loss · Ensemble Learning · Grad-CAM

1 Introduction

Agriculture poses a fundamental cornerstone of human existence, relying on the complex web of agroecosystems teeming with diverse plant life, intricate genetics, and the harmonious interplay of environmental factors governing their growth. In the context of Bangladesh, this vital sector assumes a paramount role, constituting a formidable 17% slice of the nation's gross domestic product (GDP) and casting its supportive mantle over an estimated 84% of the populace, either directly or indirectly, for their subsistence and livelihoods [1]. Therefore, crop leaf

diseases become a significant threat to both crop health and economic prosperity, undermining the delicate balance of agricultural ecosystems and impacting human well-being [2].

In a global panorama, the toll exacted by crop leaf diseases is a staggering 40% of crop losses on a worldwide scale [3]. This statistic underscores the urgent need for proactive measures, as many crop leaf diseases show visible symptoms. Therefore, timely and accurate identification of these symptoms is crucial, marking a critical moment for effectively combating agricultural diseases. Traditionally, detecting crop leaf diseases in rural landscapes relied on expert visual inspection, a challenging method for large areas that hinders prompt recognition. This labor-intensive and subjective approach is prone to errors and biases [4]. Recent advancements in automated detection have significantly mitigated the limitations of traditional methods, simplifying early-stage crop leaf disease detection. Using state-of-the-art machine learning (ML) and deep learning (DL), researchers have developed advanced image processing and object detection techniques that can surpass human capabilities in certain cases [5, 6].

Numerous researchers diligently pursued the development of automated frameworks for crop leaf disease detection, with numerous TL architectures pre-trained on the foundational ImageNet dataset [5, 7–9]. While these frameworks demonstrated impressive effectiveness, they are not immune to limitations, suggesting considerable room for improvement. Their reliance on transfer learning (TL) without model customization led to higher misclassification rates. Previous research suffered from a lack of model explainability, failure to visualize decision-making criteria, and regions of interest post-feature extraction, thereby impeding understanding and trust in outcomes and limiting practical applicability.

Moreover, most of the researchers traditionally relied on single models for classification tasks in crop leaf disease detection, overlooking the potential benefits of ensemble learning (EL) methodologies [5, 10, 11, 9]. As we explored strategies incorporating EL, a key challenge emerged: How to effectively distribute importance among constituent models within the ensemble process, significantly impacting the overall performance and robustness. This presented a critical aspect warranting further investigation and refinement of our research.

Traditionally, model selection for ensembles prioritizes accuracy, but this approach often fails with real-world datasets, leading to misclassifications and class-specific gaps. This gap between potential and real-world effectiveness highlights the need to reevaluate selection criteria for more robust ensembles. Our research focused on designing a methodology to overcome these limitations, establishing a robust ensemble framework for crop leaf disease detection. We chose the EfficientNets architecture from the TL domain for its efficiency and adaptability to different datasets, ensuring effective handling of complex tasks.

We initiated the process by adapting the chosen EfficientNets models as base models and then fine-tuning them as they were pre-trained on the ImageNet dataset. In addition to fine-tuning, we optimized the models by incorporating customized convolutional layers with diverse kernels and filter sizes, as well as a polling layer. These additions effectively captured relevant features within the

crop leaf images. Our meticulous approach extended to identifying the most suitable version of the EfficientNet architecture for each dataset and ensuring appropriate fine-tuning. To shed light on our decision-making process, we utilized the Grad-CAM.

Furthermore, we diverged from exclusive reliance on a single methodology by implementing two distinct ensemble techniques. In the context of model selection for ensemble techniques, our focus shifted towards the introduction of a novel model selection approach, departing from the conventional emphasis on accuracy to instead prioritize misclassifications and Hamming loss. We introduced a meticulously designed algorithm for model selection, with the overarching aim of minimizing misclassifications while simultaneously maximizing accuracy. Through extensive experimentation, **Bayesian Model Averaging (BMA) emerged as the most effective choice**, attributable to its remarkable proficiency in substantially reducing misclassification rates across all datasets in our experiments.

The rest of the article is structured as follows: initiates with an exploration of the existing literature in Section 2, followed by a detailed presentation of methods in Sections 3 and 4. Section 5 subsequently analyze the achieved performances. Expanding upon these findings, Section 6 engages in a comprehensive discourse, evaluating the practical implications of the model. Section 7 conscientiously delineates the study’s limitations. Finally, Section 8 concludes the paper, and encapsulates the key takeaways and contributions of the study.

2 Literature Review

In recent years, crop leaf disease detection has evolved significantly, revealing dominant trends and critical challenges. This section provides a comprehensive overview, requiring detailed exploration through an exhaustive analysis of current research.

Moin et al. [5] demonstrated avant-garde deployment of DL, employing a convolutional neural network (CNN) alongside six well-established TL models to adeptly identify diseases in pivotal crops of Bangladesh. Despite attaining notable precision rates (96.82%, and 98.55% for corn, and potato), the investigation recognizes constraints in mitigating misclassification. In another study by Islam et al. [10] utilized an eight-layer convolutional architecture in a customized CNN model, achieving 98.23% accuracy for corn and 96.90% for potato leaf disease datasets respectively. Nonetheless, limitations emerged due to the lack of regularization techniques, potentially impacting adaptability to new data scenarios. Utilizing varied TL for feature extraction and ML for corn leaf disease classification, Syarief et al. [7] achieved notable results, including 93.5% accuracy via AlexNet with support vector machine (SVM) classification; limitations arose from a lack of custom fine-tuning layers and model generalizability.

Regarding potato leaf disease detection, VGG16 acted as a feature extractor, complemented by SVM, K-Nearest Neighbors (KNN), Logistic Regression (LR), and Neural Network classifiers [8]. Notably, VGG16 coupled with LR demonstrated peak performance, achieving 97.8% accuracy, yet the study acknowledged

limitations, including potential model robustness challenges. In another study conducted by Bonik et al. [11], a seven-stage CNN model for potato leaf disease detection via a Sequential Model demonstrated promising outcomes with 94.2% accuracy. Nonetheless, it encountered limitations, e.g., a lack of customized fine-tuning layers and relatively poor performance. Furthermore, in another research by Ajra et al. [9], examined unhealthy leaf symptoms in potato and tomato leaves using image augmentation and ResNet-50 and AlexNet as CNN feature extractors. Notably, ResNet-50 attained 97% accuracy, and AlexNet achieved 96.1%; however, limitations stemmed from augmentation’s potential unreliability.

3 Datasets Description

Our investigation focused on the two predominant crops in Bangladesh, corn and potato, sourcing their leaf images from the Kaggle repository, specifically the PlantVillage dataset [12]. In table 1, we condensed an overview of the datasets of the two crop species, while Figure 1 displayed some sample images. **The corn leaf disease dataset consisted of 14,632 images classified into four distinct categories, i.e., Corn Common Rust (CCR), Corn Gray Leaf Spot (CGLS), Corn Healthy (CH), and Corn Leaf Blight (CLB). The potato leaf disease dataset comprised 8,286 images of three distinct classes, i.e., Potato Early Blight (PEB), Potato Healthy (PH), and Potato Late Blight (PLB).**

Table 1: Details of the Bangladeshi crop leaf disease datasets

Crop	Scientific Name	Total Images	Total Class	Image Format
Corn	<i>Zea mays</i>	14,632	4	JPG
Potato	<i>Solanum tuberosum</i>	8,286	3	JPG

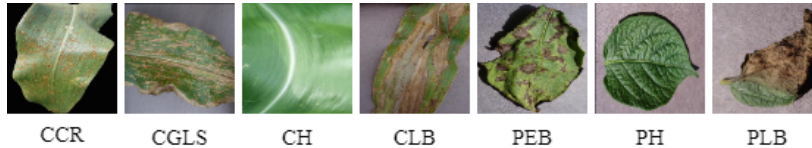


Fig. 1: Sample leaf images from Corn and Potato datasets

4 Research Methodology

Our research methodology comprised five distinct stages, starting with the application of data preprocessing techniques. Then, **we partitioned these preprocessed datasets into three subsets to seamlessly support the successive stages of training, validation, and testing** during the evaluation of our customized TL models. Subsequently, we amalgamated predictions from individual models by considering class-specific misclassifications and Hamming loss, collaboratively striving

to minimize instances of misclassification. After evaluating the results, we employed Grad-CAM to generate enlightening heatmaps, which provide profound insights into the predictive effectiveness of the model.

4.1 Data Preprocessing

The images within the Corn, and Potato leaf disease dataset had an original size of $256 \times 256 \times 3$, with the first two dimensions denoting height and width and the last dimension indicating color channels, prompting us to standardize them for our experimental setup to $224 \times 224 \times 3$.

4.2 Proposed Model Structures and Customization

Our primary emphasis revolved around the customization of existing models to adeptly leverage the principles inherent in TL paradigms. In this pursuit, **our initial step involved the meticulous construction of multiple iterations of the customized EfficientNet architectures, spanning from EfficientNetB0 to EfficientNetB7.** The preference for EfficientNet over alternative options arises from its optimized scaling and well-balanced representation of features spanning from lower to higher levels. Our thorough approach guaranteed the inclusion of fundamental and intricate attributes in our final customized models. Additionally, our model structure underwent meticulous refinement, culminating in the execution of these tailored models. Furthermore, we orchestrated a multifaceted process involving fine-tuning, parameter optimization, and architectural adaptations, ensuring the seamless integration of the models into the specific problem domain.

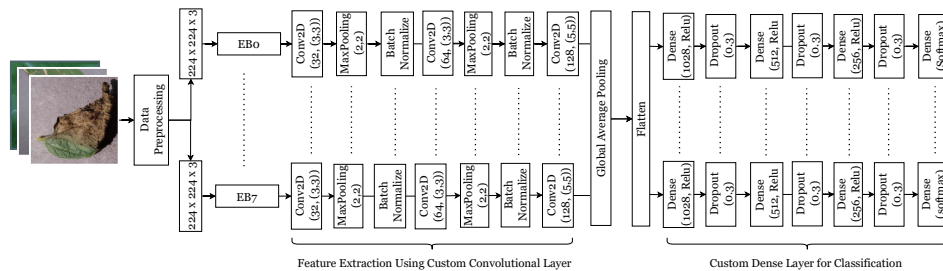


Fig. 2: Customization of EfficientNets model architecture

4.3 Customized EfficientNetB0-B7 Model Architecture

In customizing the EfficientNet (B0-B7) model architecture, we initiated by defining the base model with pre-trained weights from ImageNet, which mandated the addition of extra convolutional layers to tailor the models effectively for our specific task. **These supplementary Convolutional (Conv2d) layers were**

introduced to augment the model’s capacity for capturing intricate patterns within the data. Continuing, we added a 32-filter Conv2D layer with a 3x3 kernel from the foundational base, enhancing feature extraction and preserving spatial relationships. We followed this with MaxPooling2D to compress data, refine dimensions, and boost efficiency, all while maintaining stability through BatchNormalization (BN).

In the subsequent steps, a 64-filter Conv2D layer with a 3x3 kernel further improved feature extraction, while MaxPooling2D distilled essential traits, supported by BN for stability. Another Conv2D layer with 128 filters and a 5x5 kernel broadened receptive fields for a global context. We opted for GlobalMaxPooling2D instead of MaxPooling to generate a singular vector of maximum features and high-level abstractions. We introduced three convolutional layers, rather than just one, to facilitate the flattening of spatial hierarchies, which was crucial for effective model preparation for classification. Following this, we incorporated dense layers to enhance pattern recognition.

Moreover, we incorporated an initial dense layer consisting of 1024 neurons to learn the lower-level features, with the aid of a dropout layer to randomly drop 30% of learned neurons to prevent overfitting. Subsequently, we introduced another dense layer of 512 neurons to learn the intermediate-level features. Moreover, a third dense layer of 256 neurons was added to learn the higher-level or complex features. The final and last dense layer consisted of the number of neurons related to class values and activated with a softmax activation function. The decision to include three dense layers, as opposed to just one, aimed to equip the model with the capacity to effectively learn and represent diverse patterns for various crop leaf disease detection. The architecture seamlessly integrates fine-tuning with customized layers, as exemplified in Figure 2, representing the highest level of expertise in crop ailment through careful adaptation of the EfficientNet framework.

4.4 HAMIP Ensemble Techniques

We embraced a systematic strategy known as Hamming loss and Misclassification Inverse Proportionate (HAMIP) to elevate our ensemble’s precision to attain peak performance. This approach intricately guided the curation of models from our collection of 8 meticulously customized EfficientNet models. Our strategy

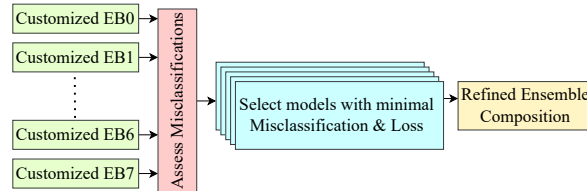


Fig. 3: Illustration of refined ensemble formation through model curation

aimed to systematically reduce class-specific misclassifications by selecting models with the fewest errors and, when comparable, prioritizing those with lower Hamming loss, thereby enhancing ensemble precision, resilience, and predictive efficacy. Refer to Figure 3 for visual representation, and the procedural steps for the implementation of HAMIP are as follows:

Class-Specific Misclassification Computation In the initial step, we systematically computed class-specific misclassification values for all classifiers present in the ensemble. This process involved evaluating the performance of each classifier concerning distinct classes. We determined the misclassification $M_{c,i}$ for a given class c and classifier i using the equation:

$$M_{c,i} = \sum_{j \neq c} ConfusionMatrix_{ij} \quad (1)$$

Here, the summation encompasses all instances j where the predicted class j does not match the true class c .

Selective Classifier Identification Subsequently, we entailed the identification of classifiers that demonstrate the most minimal misclassification rates across distinct classes. This selection process ensured that each class’s classifier was chosen, aligning to achieve the lowest $M_{c,i}$ values. This meticulous classifier selection guaranteed precision in predicting the class-wise outcomes of our proposed model.

Hamming Loss Assessment In scenarios where multiple models exhibit similar misclassification, our preference shifted towards the lower Hamming loss (HL) models. This criterion guarantees the selection of models that reduce errors specific to individual classes and demonstrate improved accuracy in labeling. The selection process incorporates the Hamming loss HL_i for each model i , defined as:

$$HL_i = \frac{1}{N} \sum_{j=1}^N \delta(y_{ij}, \hat{y}_{ij}) \quad (2)$$

Here, N is denoted as the total number of instances, and for instance, j and class i , y_{ij} represent the true label, and \hat{y}_{ij} as the predicted label for the same instance, and δ is the Kronecker delta function, which equals 1 if $y_{ij} \neq \hat{y}_{ij}$ (where ij are indices) and 0 otherwise.

Dual Criterion for Ensemble Refinement In this phase, we seamlessly integrated the preferred optimal classifiers for each class, then employed diverse ensemble techniques and placed paramount importance on their performance concerning misclassification and Hamming loss. Here’s our ultimate algorithm for selecting the models with the least misclassification and Hamming loss for each class in a dataset:

Algorithm 1 Model Selection based on Misclassification and Hamming Loss

```

1: procedure SELECTMODELS(dataset)
2:    $classes \leftarrow$  List of classes in the dataset
3:    $selected\_models \leftarrow \emptyset$ 
4:   for  $class$  in  $classes$  do
5:      $least\_misclassified\_models \leftarrow$  Models with the least misclassification for
        $class$ 
6:     if  $|least\_misclassified\_models| > 1$  then
7:        $selected\_model \leftarrow$  Model with the least Hamming loss among
        $least\_misclassified\_models$ 
8:       for  $model$  in  $least\_misclassified\_models$  do
9:         if  $model$  has lower Hamming loss than  $selected\_model$  then
10:          Add  $model$  to  $selected\_models$ 
11:        end if
12:      end for
13:     else if  $|classes| > |models|$  then
14:       Allows taking one model for multiple classes
15:     else
16:       Add  $least\_misclassified\_models$  to  $selected\_models$ 
17:     end if
18:   end for
19:   return  $selected\_models$ 
20: end procedure

```

4.5 Ensemble Learning (EL) Strategies

We utilized four ensemble techniques Majority Voting (MV), SoftMax Averaging (SMA), Weighted Averaging (WA), and Bayesian Model Averaging (BMA) to evaluate which one performs better.

4.6 Performance Evaluation Measures

To comprehensively assess our models' performance across diverse datasets, we used confusion matrices to derive various statistical indicators, providing a holistic evaluation of the model's efficacy. We incorporate a set of performance metrics, including class-wise misclassifications ($M_{c,i}$), which have been previously detailed in HAMIP techniques. Additionally, we measure total misclassifications overall (MT), accuracy, precision, recall, specificity, and F1-score to serve as a comprehensive evaluation framework for the overall effectiveness of our model.

4.7 Experimental setup

We utilized Kaggle's computational resources, we employed a 2-core Intel Xeon CPU and a high-performance Tesla P100 GPU for swift training and inference. We carefully divided the dataset into training (70%), validation (15%), and testing (15%) subsets to ensure robust evaluation. With a precise batch size of 16,

we utilized the Adam optimizer with a learning rate of 0.001 and an epsilon value of 0.1 for model convergence. The training involved the categorical cross-entropy loss function over 40 epochs, with an early stopping checkpoint after 15 consecutive epochs without improvement in validation accuracy, ensuring optimized results and resource utilization in our thorough training and evaluation setup.

5 Experimental Results Analysis

In this section, we unveil the empirical and graphical outcomes stemming from our meticulous evaluation of models across two distinct leaf disease datasets. A rigorous comparison was conducted, evaluating the performance of customized EfficientNet models ranging from B0 to B7, with model selection based on misclassification and Hamming loss criteria. We then assessed four tailored ensemble techniques for corn, and potato datasets, conducting a comprehensive analysis with various experimental measurements and performance metrics to evaluate the model’s effectiveness in addressing crop leaf diseases.

In our pursuit of corn leaf disease detection, we initially employed eight customized pre-trained models, spanning from EfficientNet B0-B7. The evaluation of their performance across multiple dimensions, such as misclassification for each class, is presented in Table 2, while the corresponding accuracy, Hamming loss (HL), precision, recall, F1 score, and specificity are displayed in Table 3.

Table 2: Misclassifications for corn dataset’s classes

Class	EB0	EB1	EB2	EB3	EB4	EB5	EB6	EB7
CCR	7	4	1	1	1	0	0	0
CGLS	6	7	6	9	6	7	7	10
CH	0	0	0	1	0	0	0	0
CLB	12	5	5	11	3	8	7	9
Total	25	16	12	22	10	15	14	19

Remarkably, we observed that EfficientNetB4 (EB4) emerged as the top gainer, boasting an impressive accuracy of 99.55% and a minimal 0.45% Hamming Loss. Furthermore, EB4 exhibited the highest precision, recall, F1 score, and specificity. Subsequently, we applied our customized dual-criterion process for model selection from the EB0-EB7 range. Analyzing Tables 2 and 3, we identified that, for the four-class corn leaf dataset, class CCR exhibited zero misclassifications, with EB5, EB6, and EB7 showcasing exemplary performance. Among these, we selected EB6 as the model for the ensemble due to its lowest HL. Furthermore, for the category CGLS, we maintained EB4, while for class CH, we introduced EB2 as the designated model. Notably, EB5 displayed a lower HL than EB2, warranting its inclusion in our ensemble selection.

After the selection of our models, we proceeded to apply four ensemble techniques, among which BMA delivered better performance. Additionally, we compared our model selection methods and ensemble techniques with accuracy-based

Table 3: Evaluation of EB0-EB7 performance on corn leaf disease dataset

Algorithm	Accuracy(%)	HL(%)	Precision(%)	Recall(%)	F1-Score(%)
EB0	98.86	1.14	98.87	98.86	98.86
EB1	99.27	0.73	99.27	99.27	99.27
EB2	99.45	0.55	99.45	99.54	99.45
EB3	98.95	1.05	98.95	98.95	98.95
EB4	99.55	0.45	99.54	99.45	99.54
EB5	99.32	0.68	99.32	99.31	99.32
EB6	99.36	0.64	99.36	99.36	99.36
EB7	99.14	0.86	99.16	99.15	99.13

BMA (Acc_BMA). From the data presented in Table 3, it became evident that EfficientNet models EB2, EB4, and EB6 emerged as the top three accuracy gainers, and thus, we also employed Acc_BMA for their ensembling to facilitate comparative analysis, as detail is presented in Table 4. It became evident that BMA with our model selection method outshone the rest in terms of performance, with only one misclassified sample and leading across all performance metrics. Figure 4 illustrates the confusion matrix and ROC curve of the top-performing BMA model, affirming the effectiveness of our overall approach. During our in-depth

Table 4: Evaluation of ensemble methods on corn leaf disease dataset

Measures	MV	SMA	WA	BMA	Acc_BMA
Accuracy(%)	99.90	99.90	99.77	99.95	99.81
HL(%)	00.09	00.09	00.23	00.04	00.18
MT	2	2	5	1	4
Precision(%)	99.90	99.90	99.77	99.95	99.81
Recall(%)	99.90	99.90	99.77	99.95	99.81
F1-Score(%)	99.90	99.90	99.77	99.95	99.81
Specificity(%)	99.97	99.96	99.92	99.98	99.93

Table 5: Misclassifications for Potato leaf disease dataset's Classes

Class	EB0	EB1	EB2	EB3	EB4	EB5	EB6	EB7
PEB	0	5	3	1	2	3	2	1
PH	0	3	0	4	1	0	1	0
PLB	3	1	5	4	13	2	5	1
Total	3	9	8	9	16	5	8	2

assessment of the customized EfficientNet models (EB0-EB7) specifically for the potato leaf disease dataset, we undertook a thorough and exhaustive evaluation. The misclassifications for each class are meticulously documented in Table 5, while additional performance metrics analysis is in Table 6. It's worth highlighting that EfficientNet B7 (EB7) exhibited outstanding performance, achieving an impressive 99.84% accuracy, the highest precision at 99.84%, and remarkable

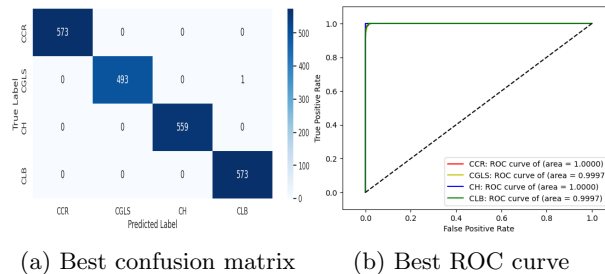


Fig. 4: Confusion matrix and ROC curve for corn leaf disease dataset

Recall and F1 scores of 99.83%. Equally noteworthy, EB0 also delivered commendable results, with just 3 instances of misclassification. Next, after reviewing Table 5 and Table 6 and applying our customized algorithm, we observed that in the three-class potato leaf disease dataset, class PEB had zero misclassifications, with EB0 as the solitary top performer, and we designated EB0 as the chosen model.

Table 6: Evaluation of EB0-EB7 performance on potato leaf disease dataset

Algorithm	Accuracy(%)	HL(%)	Precision(%)	Recall(%)	F1-Score(%)
EB0	99.76	0.24	99.76	99.75	99.75
EB1	99.28	0.72	99.28	99.27	99.28
EB2	99.36	0.64	99.36	99.35	99.36
EB3	99.28	1.04	99.28	99.27	99.27
EB4	98.72	1.28	98.73	98.71	98.71
EB5	99.60	0.40	99.60	99.59	99.60
EB6	99.36	0.64	99.36	99.36	99.36
EB7	99.84	0.16	99.84	99.83	99.83

Regarding class PH, EB0, EB2, EB5, and EB7 exhibited the lowest misclassification, with EB7 emerging as the preferred choice due to its minimal HL. Likewise, for class PLB, EB1 and EB7 had the fewest misclassifications. Since EB7 had already been selected, we opted for EB1. Notably, EB5 and EB6 demonstrated lower HL values than EB1, prompting their inclusion in our ensemble selection. To compare our selection method with accuracy-based selection, we identified EB0, EB5, and EB7 as the top accuracy gainers from Table 6 and integrated them to calculate Acc_BMA. After that, the results of four ensemble techniques along with Acc_BMA for potato leaf disease detection are shown in Table 7. BMA continued to excel, exhibiting zero misclassifications and achieving a flawless 100% performance across all performance measuring metrics. Figure 5 serves as additional validation for our approach, showcasing the confusion matrix and ROC curve of the high-performing BMA model. The exceptional results achieved in potato leaf disease detection can be primarily attributed to the substantial contributions made by our tailored pre-trained models and the ensemble model selection algorithm.

Table 7: Evaluation of ensemble methods on potato leaf disease dataset

Measures	MV	SMA	WA	BMA	Acc_BMA
Accuracy(%)	99.91	99.91	99.91	100.00	99.91
HL(%)	00.08	00.08	00.23	0.00	00.08
MT	1	1	1	0	1
Precision(%)	99.91	99.91	99.92	100.00	99.91
Recall(%)	99.91	99.91	99.91	100.00	99.91
F1-Score(%)	99.91	99.91	99.91	100.00	99.91
Specificity(%)	99.95	99.95	99.97	100.00	99.95

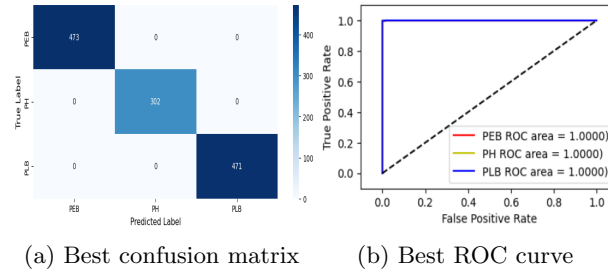


Fig. 5: Confusion matrix and ROC curve for Potato leaf disease dataset

5.1 Gradient Class Activation Map (Grad-CAM)

In our study, we incorporated Grad-CAM visualization, a technique enhancing our understanding of focal model regions during prediction. This visual representation, illustrated in Figure 6, notably enhances transparency and clarity surrounding our best-performing base model’s internal operations.

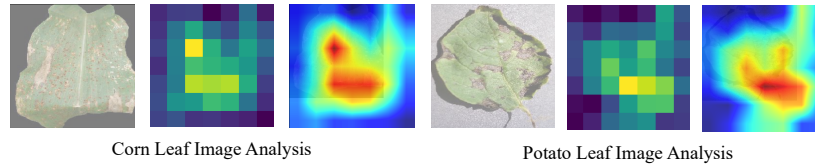


Fig. 6: Gradient class activation map visualization

Firstly, we generated a heatmap by mapping the input image to the final convolutional layer activations and output predictions. Next, we calculated the gradient of the top predicted class concerning the last convolutional layer’s activations. We assigned weights to each channel based on its significance in predicting the class and aggregated them for the class activation heatmap. To facilitate visualization, we standardized the heatmap within a scale of 0 to 1. The final step involves rescaling the heatmap to a range of 0-255 and visualizing it, providing valuable insights into our model’s decision-making process.

6 Discussion and Extended Comparison

Our carefully crafted approach, thoughtfully tailored to the intricacies of crop leaf disease detection, employed diverse datasets encompassing corn, and potato. Our strategy introduced customized TL models rooted in the foundation of EfficientNet B0-B7 architecture while innovatively selecting ensemble models based on misclassification and Hamming loss. Our evaluation highlighted EB4’s excellence in corn and EB7’s superiority in potato ultimately leading to a refined ensemble selection process.

In a rigorous analysis of our work, we found that BMA consistently emerged as the superior ensemble technique, remarkably reducing misclassified images. It’s important to note that we also incorporated accuracy-based selection into our approach, choosing the top three models based on accuracy and then ensemble them using Acc_BMA. The utilization of our model-selected BMA approach yielded superior results. In the case of potatoes, our approach achieved a remarkable absence of misclassifications, while in corn, only two instances were misclassified. These outcomes underscore the impressive efficacy of our proposed BMA model in enhancing classification accuracy, primarily due to its capacity to capture diverse perspectives from constituent models and minimize the impact of individual model biases.

In addition, we complemented our evaluation with GradCAM visualizations, providing insightful visual representations of our best-performing base models, namely, EB4 for corn and EB7 for potato, further enhancing the interpretability and trustworthiness of our results. Our holistic model proposal integrated customized EfficientNet models, a tailored selection algorithm, and the potent BMA technique, reinforced by insightful GradCAM visualizations, constituting a robust framework for crop leaf disease detection.

Table 8: Performance comparison with previous researches

Dataset	Reference	Accuracy(%)	Precision(%)	Recall(%)	F1-Score(%)
Corn	[5]	96.82	-	-	-
	[10]	98.23	-	-	-
	[7]	93.0	-	93.5	-
	Proposed	99.95	99.95	99.95	99.95
Potato	[10]	96.90	-	-	-
	[8]	97.8	97.8	97.8	97.8
	[11]	94.40	73.00	80.00	82.30
	Proposed	100.00	100.00	100.00	100.00

Furthermore, we performed an exhaustive comparison which is outlined in Table 8 firmly establishes the superiority of our proposed methodology over previous research efforts. This achievement holds particular significance, given that our approach prioritizes interpretability without compromising the depth and comprehensiveness of the framework.

7 Threats To Validity

While we diligently tailored our crop leaf disease prediction framework to two specific crop varieties (corn, and potato), our focused approach may have limited the generalizability of our findings to these crops. It was essential for us to acknowledge that benchmark datasets for crop leaf disease detection encompassed other prominent species like wheat, tomato, rice, and chili leaf diseases, presenting a potential avenue for future research. The prospect of merging diverse leaf disease datasets across various crop species could have expanded the dataset’s scope, enabling us to develop comprehensive prediction models for each distinct crop leaf disease, a direction we found enticing for further exploration.

In this work, we primarily concentrated on presenting techniques and a proposed method for leaf disease detection, yielding minimal misclassifications. We postulated that integrating attention mechanisms, such as soft attention (SA), could further enhance results by focusing on subtle yet critical features that might be challenging to detect in the remaining misclassifications. We aim to incorporate these enhancements in future extensions of this research, advancing the sophistication and performance of our predictive models.

8 Conclusion and Future Work

Our work, precisely designed to address the intricate landscape of crop leaf disease detection, introduces a strategic response that combines innovative techniques and a tailored model. Utilizing diverse datasets encompassing corn, and potato, we embarked on a journey of exploration, beginning with customized TL and an exhaustive examination of various EfficientNet versions. Our research fills a critical void in ensemble learning within the realm of DL methodologies.

In this research, we challenged the conventional model selection paradigm that primarily relies on accuracy. Recognizing that accuracy alone may not adequately address misclassifications and gaps within each class, we introduced a novel approach that factors in misclassification counts and Hamming loss, revolutionizing ensemble model selection. The remarkable results achieved through BMA highlight the efficacy of our approach. We also identified the best-performing EfficientNet models for each dataset, showcasing EB4 for corn and EB7 for potato. Additionally, the visualization of their decision-making process using GradCAM provided valuable insights. Our primary aim was to minimize misclassifications, with notable success in the potato leaf dataset and substantial progress in the corn dataset.

In the future, we intend to extend our research to encompass diverse crops, exploring new algorithms for ensemble model selection. Further investigations into different attention mechanisms are on the horizon, aiming to create a model with minimal misclassifications, approaching near-zero levels.

References

1. Rahman, M.T.: Role of agriculture in bangladesh economy: uncovering the problems and challenges. *Intl. J. of Business and Management Invention* **6**(7) (2017)
2. Jannat, N., Hasan, S.M., Srizon, A.Y., Oishe, M., Efat, A.H., Taraque, M.F., Mitu, M., Faruk, M.F.: Efficient detection of crop leaf diseases: A lightweight convolutional neural network approach for enhanced agricultural productivity. In: *2023 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*. pp. 99–103. IEEE (2023)
3. Han, L., Haleem, M.S., Taylor, M.: A novel computer vision-based approach to automatic detection and severity assessment of crop diseases. In: *2015 Science and Information Conference (SAI)*. pp. 638–644. IEEE (2015)
4. Bock, C., Poole, G., Parker, P., Gottwald, T.: Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. *Critical reviews in plant sciences* **29**(2), 59–107 (2010)
5. Moin, N.B., Islam, N., Sultana, S., Chhoa, L.A., Howlader, S.R.K., Ripon, S.H.: Disease detection of bangladeshi crops using image processing and deep learning—a comparative analysis. In: *2022 2nd International Conference on Intelligent Technologies (CONIT)*. pp. 1–8. IEEE (2022)
6. Ngugi, H.N., Ezugwu, A.E., Akinyelu, A.A., Abualigah, L.: Revolutionizing crop disease detection with computational deep learning: a comprehensive review. *Environmental Monitoring and Assessment* **196**(3), 302 (2024)
7. Syarief, M., Setiawan, W.: Convolutional neural network for maize leaf disease image classification. *Telkomnika (Telecommunication Computing Electronics and Control)* **18**(3), 1376–1381 (2020)
8. Tiwari, D., Ashish, M., Gangwar, N., Sharma, A., Patel, S., Bhardwaj, S.: Potato leaf diseases detection using deep learning. In: *2020 4th international conference on intelligent computing and control systems (ICICCS)*. pp. 461–466. IEEE (2020)
9. Ajra, H., Nahar, M.K., Sarkar, L., Islam, M.S.: Disease detection of plant leaf using image processing and cnn with preventive measures. In: *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*. pp. 1–6. IEEE (2020)
10. Islam, M.T.: Plant disease detection using cnn model and image processing. *Int. J. Eng. Res. Technol* **9**(10), 291–297 (2020)
11. Bonik, C.C., Akter, F., Rashid, M.H., Sattar, A.: A convolutional neural network based potato leaf diseases detection using sequential model. In: *2023 International*
12. *Bangladeshi crops disease dataset - kaggle* (verified march 2024). <https://www.kaggle.com/datasets/nafishamoin/bangladeshi-crops-disease-dataset>
13. Upadhyay, D., Manero, J., Zaman, M., Sampalli, S.: Intrusion detection in scada based power grids: Recursive feature elimination model with majority vote ensemble algorithm. *IEEE Trans. on Network Science and Eng.* **8**(3), 2559–2574 (2021)
14. Nguyen, T., Pernkopf, F., Kosmider, M.: Acoustic scene classification for mismatched recording devices using heated-up softmax and spectrum correction. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp. 126–130. IEEE (2020)
15. Solano, E.S., Affonso, C.M.: Solar irradiation forecasting using ensemble voting based on machine learning algorithms. *Sustainability* **15**(10), 7943 (2023)
16. Wang, G., Jia, R., Liu, J., Zhang, H.: A hybrid wind power forecasting approach based on bayesian model averaging and ensemble learning. *Renewable energy* **145**, 2426–2434 (2020)