



# Automatic Early Classification of Cassava Leaf Disease with Ensemble of Lightweight Models

Kaushik Vishwakarma Udaybhan Rathore Rupesh Kumar Amit Shakya  
Shruti Phutke Chetan Gupta Lalit Sharma

This paper, titled “Automatic Early Classification of Cassava Leaf Disease with Ensemble of Lightweight Models,” was presented at CVR-2025 (5th International Conference on Computer Vision and Robotics), held in hybrid mode at the National Institute of Technology Goa, India, from April 25–26, 2025.

Reprinted with permission. Copyright © 2025 CVR and SCRS.

Further use or distribution is not permitted without permission from CVR and SCRS.

## Abstract

The growing demand for healthy food, driven by population growth and the prevalence of plant diseases, is a significant concern. Farmers are increasingly forced to cultivate their fields continuously, often relying on pesticides that degrade soil quality and contribute to the spread of various diseases. Disruptive technologies, such as automated cassava leaf disease detection using deep learning models, can play a vital role in promoting agricultural sustainability. While numerous deep learning methods have been explored, challenges remain, particularly in developing lightweight models that can quickly and accurately identify disease class variations while also being suitable for deployment on electronic devices. This research focuses on evaluating lightweight deep learning models, specifically CNN and transformer networks, for cassava leaf disease detection. Based on our experiments, we propose an ensemble model consisting of lightweight models (CNN and transformer) for automated early disease detection using raw images. We have utilized the ResNeXt, EfficientNet-B5, and TinyViT lightweight models, optimizing the ensemble process through brute force approaches. Additionally, we experimented with centre-crop and multi-crop image transformations to test these models, aiming to enhance performance and classify images based on the global information derived from the whole image. Our combined approach achieved state-of-the-art results with an overall recall of 90.35% on the unseen test cassava leaf disease dataset. With high accuracy, fewer parameters, and low computation time, our model is well-suited for deployment on mobile devices.

## 1

## INTRODUCTION

Cassava is a widely cultivated staple crop in sub-Saharan Africa, providing essential nutrition and energy. While the leaves and roots are both suitable for consumption, the roots are the preferred choice for their high energy content and versatility—they can be eaten raw, roasted, boiled, or processed in various ways<sup>[1][2]</sup>. The leaves and tender shoots are rich in protein and vitamins and are used as vegetables in many regions. Farmers across Africa grow cassava at different scales and under diverse geographical and weather conditions, contributing significantly to food stability and industrial development. Although widely cultivated, cassava is highly prone to

various diseases, including infections caused by lesser-known viral viruses. The four most common diseases, each with distinct signs and symptoms, are listed below and illustrated in Figure 1.

**Cassava Bacterial Blight (CBB):** CBB is one of the most devastating diseases, causing the highest yield losses among cassava crops worldwide. In the 1970s, it severely impacted Africa, with countries like Nigeria, Uganda, and Zaire suffering crop losses of 75% to 100%, leading to malnutrition and starvation in some regions. The disease thrives in moist conditions, affecting cassava plants in such areas. Symptoms include black leaf spots and blights, with affected leaves wilting, drying prematurely,

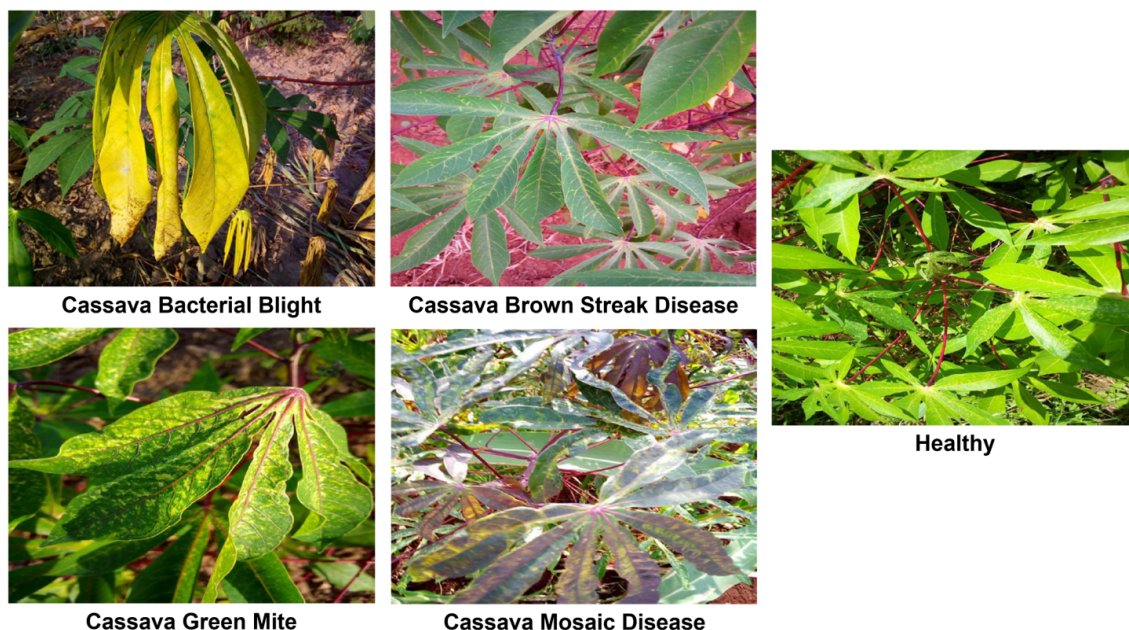


Fig. 1 Provide sample images of the four major cassava leaf diseases together with healthy cassava leaf sample image from the dataset

and eventually shedding off.

**Cassava Green Mite (CGM):** CGM is a pest originating from South America, was introduced to Africa in the early 1970s and has now spread to 27 countries. It creates white spots on cassava leaves, beginning with small lesions that progressively expand, covering the entire leaf. This leads to chlorophyll loss and hinders photosynthesis. In cases of severe infestation, mottling symptoms may appear similar to those of cassava mosaic disease, making differentiation between the two difficult. Infested leaves commonly lose moisture, shrink, and break away from the plant, contributing to further damage.

**Cassava Brown Streak Disease (CBSD):** CBSD, which emerged in Eastern Africa in the 1930s, remains a major threat to cassava production, causing root yield losses of up to 70%. It is transmitted by whiteflies and is characterized by yellowing along the veins, which may expand into larger yellow patches. CBSD severely impacts tuber roots, forming dark-brown necrotic lesions that reduce root size and degrade the quality of the crop.

**Cassava Mosaic Disease (CMD):** CMD was first reported in 1894 by War-burg in Tanzania. It remains a significant

constraint to cassava production in Africa, contributing to famine and, in some regions, even death. The disease typically causes foliar symptoms like mottling, mosaic patterns, leaf deformation, and a significant decrease in both leaf and plant size. Affected leaves often display green patches mixed with yellow or white discoloration, diminishing the surface area for photosynthesis and resulting in poor growth and crop yields.

Distinguishing between different cassava diseases is challenging and often requires farmers' assistance, which is both time-consuming and resource-intensive<sup>[3]</sup>. The inability to detect infections early can lead to significant economic losses, highlighting the need for automated approaches to assist with timely disease detection and prevention. While several machine learning models have been explored in this area, recent deep learning models have delivered state-of-the-art (SOTA) results. However, these models typically require a large number of parameters, increasing their computational cost, and making them unsuitable for resource-constrained devices like mobile phones. In our research, we focused on lightweight deep-learning models that use fewer parameters while maintaining high accuracy and efficiency. We experimented with two architectures,

namely, convolutional neural networks (CNN) and transformers, and proposed an ensemble model that combines the strengths of both types of networks in lightweight constrained. During testing, the input image is cropped at the centre for specific model (CNNs/transformer) input requirements. To mitigate information loss caused by centre-cropping when passing images through the model, a multi-crop approach is utilized. In this, each image is divided into five overlapping crops based on the model's input size, ensuring the entire image is covered. Each crop was then passed through the model to predict individual scores, which were subsequently merged to produce the final disease classification. Collectively, our ensemble model, combined with this image transformation approach, achieved an overall recall of 90.35% and an F1-score of 0.902 on the test dataset, setting a new state-of-the-art with fewer parameters and reduced computation time.

The primary contributions of this work are as follows:

- Conducted experiments on various lightweight models using two types of architectural, CNN-based and Transformer-based networks, for multi-class classification tasks.
- Analysis the performance of these models under two distinct testing strategies on the test datasets, namely centre-crop transformation and multi-crop transformation.
- We propose a novel ensemble model optimized using a brute-force algorithm, combined with multi-crop transformation testing strategies, which achieved state-of-the-art results on unseen test datasets.

## 2

## LITERATURE REVIEW

In artificial intelligence, two prominent types of vision-based frameworks—machine learning (ML) and deep learning (DL), are utilized for plant disease recognition. While ML-based models are generally lighter and less computationally intensive than DL-based models, their recognition accuracy often lags behind that of DL models<sup>[4]</sup>. Consequently, recent research has increasingly

shifted toward DL-based approaches, particularly for plant disease recognition. Among deep learning methods, Convolutional Neural Networks (CNNs) have demonstrated promising results in plant disease detection. This is evident in comparative studies evaluating various CNN architectures, such as VGGNet ResNet50<sup>[6]</sup>, DenseNet-BC-121-32<sup>[7]</sup>, and InceptionV3<sup>[8]</sup>, for cassava leaf disease classification. Among these architectures, DenseNet-BC-121-32 achieved the highest accuracy of 80.52%, followed by InceptionV3 at 77.25%, ResNet50 at 75.76%, while VGGNet exhibited comparatively lower performance. Further studies highlight the impact of dataset imbalance on model performance. For instance, experiments conducted on an imbalanced cassava disease dataset from Kaggle<sup>[9]</sup> demonstrated a 10% improvement in accuracy when employing Pulse- Coupled Neural Networks (PCNN) and Deep Residual Neural Networks (DRNN) compared to conventional CNN-based approaches. To enhance model robustness, the CNN-based Plant Disease Detection (PDD) framework incorporates data augmentation approaches and applies multilevel and multiscale feature extraction to establish a class- and scale-invariant architecture<sup>[10]</sup>. The PDD-Net model surpasses baseline architectures, achieving an overall of recall of 85.77%, precision of 86.41%, F1-score of 86.02%, and an accuracy of 86.98% on the cassava leaf disease dataset. Despite their success, CNN-based approaches are often computationally expensive and require substantial parameters, posing challenges for deployment on edge devices<sup>[11]</sup>. To address these limitations, an Enhanced Convolutional Neural Network (ECNN) was proposed for real-time cassava leaf disease identification<sup>[12]</sup>. The ECNN employs depth-wise separable convolutions to reduce both computational complexity and feature redundancy while maintaining classification accuracy.

The cassava dataset presents several challenges for training CNN models, primarily due to its small size and class imbalance, with certain disease classes being overrepresented. To address these issues and improve performance, a comparative study<sup>[13]</sup> is conducted to evaluate the Hybrid Ensemble Disease Detection Model

(HEDTM) against custom CNN architectures. This ensemble model, which integrates InceptionV3, Xception, and DenseNet-BC-121-32, achieved the highest accuracy of 88.83%, surpassing other ensemble configurations such as ResNet50V2 + DenseNet-BC-121-32 and ResNet50 + ResNet50V2. The results of this study highlight the potential of leveraging the strengths of individual models within a unified ensemble approach. Recently, researchers have also been attracted toward vision transformer-based models to address the limitations of CNN-based approaches. In the study by<sup>[14]</sup>, the Vision Transformer (ViT) was employed instead of traditional CNNs for classifying cassava leaf diseases. Experimental results demonstrated that ViT achieved at least 1% higher accuracy than popular CNN models like EfficientNet and ResNet50d on the Cassava Leaf Disease Dataset, highlighting its potential superiority in leaf disease analysis. Building on ViT, proposed a deep learning method for identifying viral diseases in cassava leaves, achieving a classification accuracy of 90.02% on the private test set after applying K-Fold cross-validation<sup>[15]</sup>. Training transformer models demands substantial computational resources and large datasets. While they achieve strong results, their deployment on mobile devices remains challenging. Consequently,

lightweight CNNs have gained traction in plant disease detection due to their efficiency on resource-constrained devices like smartphones and edge devices. For example, MobileNet-V2 enhanced using a class activation map to achieve better cassava leaf disease classification<sup>[16]</sup>.

Thus, our objective is to propose a novel model that integrates the strengths of both the networks, CNN and transformer-based to achieve superior classification performance with reduced computational complexity, making it suitable for mobile and edge applications through ensemble learning.

### 3

## EXPERIMENTAL SETUP

### 3-1. Dataset

This dataset was initially released for the Kaggle Cassava Leaf Disease Classification Challenge<sup>[17]</sup>. It contains a total of 21,397 training images, divided into five categories: CBB (1,087 images), CGM (2,386 images), CBSD (2,189 images), CMD (13,158 images), and Healthy (2,577 images). The dataset is allocated into 80% for training and 20% for validation, as illustrated in Figure 2. In the data processing step, the images were cropped to  $512 \times 512$  and normalized to minimize background noise and

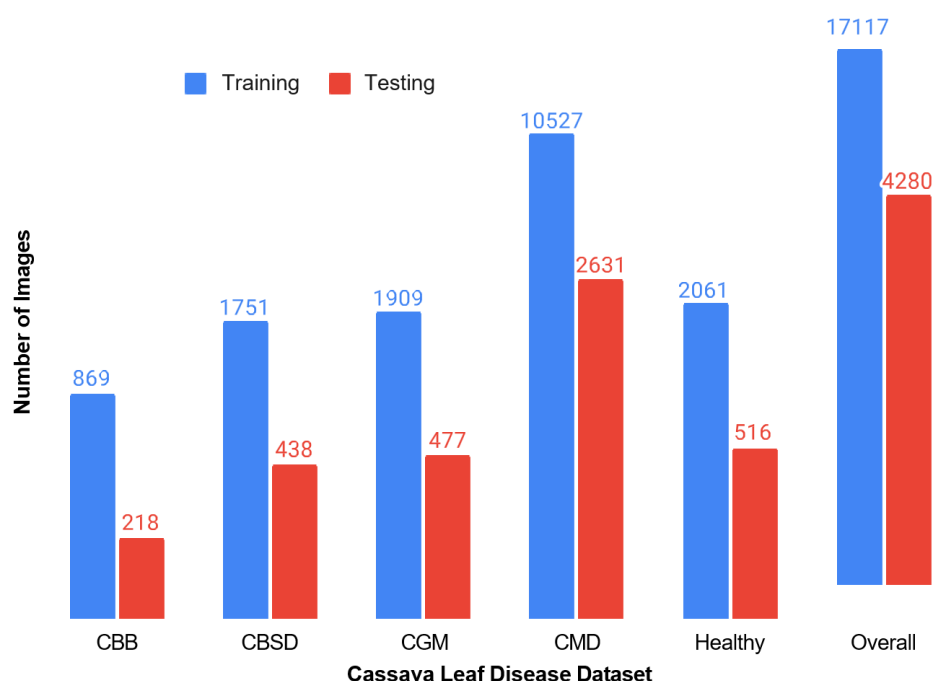


Fig. 2 The bar chart displays the distribution cassava leaf disease classification dataset

focus on critical features relevant to classification in training and testing. For model training, various transformations are applied to the training data to ensure generalized and stable model training, including transposition, horizontal flip, vertical flip, shift-scale rotation, and normalization. Since resizing from  $800 \times 600$  to  $512 \times 512$  can sometimes remove infected parts of the image affected by the disease. Therefore, we adopted a technique that was used to create five  $512 \times 512$  crops from each original image. To ensure the model had a comprehensive view for accurate disease detection during testing, five crops were taken from each image: four from the corners and one from the centre, as illustrated in Figure 3. The final prediction for each image was generated by averaging the class-wise probabilities across all five cropped regions. We evaluated the model's performance using both cropping strategies, which we refer to as centre-crop and multi-crop. Furthermore, to maximize the use of the training data, we implemented K-fold cross-validation using a 5-fold approach, dividing the dataset into five equal subsets. In each iteration, one subset was designated for validation, while the remaining four were used for training. This process was repeated five times, ensuring each subset served as the validation set once and as part of the training set four times. This method provides a comprehensive evaluation of each model's performance and generalization ability.

### 3-2. Models

In this research, we experimented with a lightweight deep-learning architecture designed to use fewer parameters while achieving higher accuracy in classifying cassava leaf diseases. Specifically, we utilized two types of models: CNN-based and transformer-based architecture, as detailed below:

**ResNeXt<sup>[18]</sup>**: ResNeXt, short for residual networks with external transformations, is a CNN-based architecture that builds the foundational principles of the ResNet (Residual Network) model. Unlike ResNet, which relies on multiple smaller paths, ResNeXt introduces groups of parallel paths. Each group contains multiple paths, with each path learning different features. This structured grouping allows the network to capture a broader range of features more effectively, eventually enhancing its overall classification of power. We fine-tuned ResNeXt50 using two distinct approaches. In the first approach, we directly added a classification layer, resulting in a model we refer to as ResNeXt50 23M. This model contains 22,990,149 parameters (approximately 23 million) and requires 90 MB of storage. In the second approach, we extended the fine-tuning by incorporating additional layers between ResNeXt50 and the final classification layer to facilitate more gradual transfer learning. This extended model, named ResNeXt50 26M, comprises 25,676,613 parameters (approximately 26 million) and occupies 100

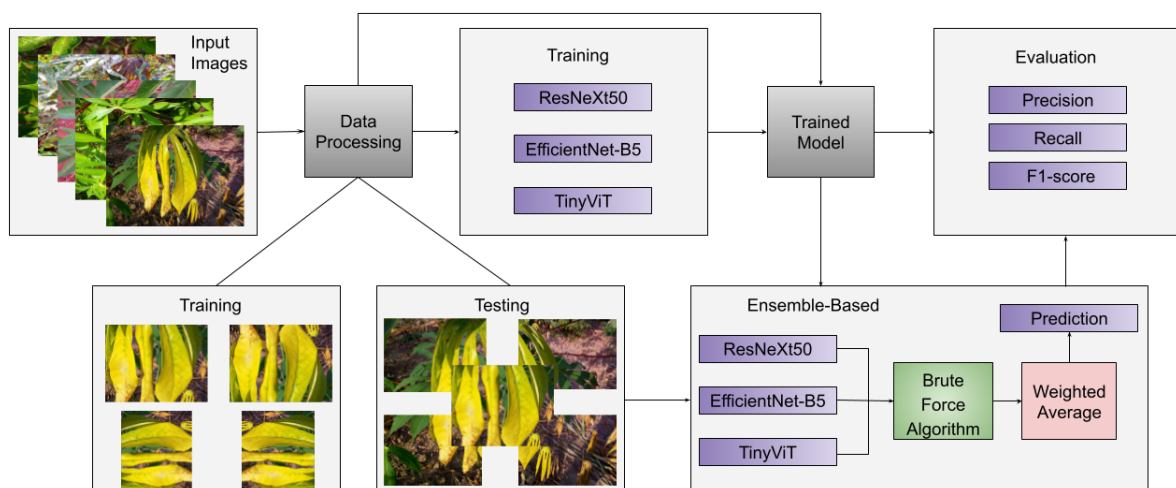


Fig. 3 Pipeline of the proposed ensemble model with a unique image transformations technique for training and testing images

MB of storage. These models were employed to evaluate the impact of parameter size on classification performance and efficiency.

**EfficientNet-B5<sup>[19]</sup>:** EfficientNet is a CNN architecture that uses a compound scaling method to proportionally adjust width, depth, and resolution, achieving both high accuracy and computational efficiency. The EfficientNet family provides a spectrum of models that balance complexity and accuracy, achieving strong classification performance with fewer parameters and reduced computational demands, while maintaining efficiency. Similarly, we utilized two variants of the EfficientNet B5 model to evaluate the impact of parameter size on inference time and overall efficiency. The first variant, similar to the fine-tuning approach of ResNeXt50 23M, is named EfficientNetB5 28M. It contains 28,351,029 parameters (approximately 28 million) and occupies 108.15 MB of storage. The second variant, similar to ResNeXt50 26M, is an extended version named EfficientNet-B5 31M. It features 31,037,493 parameters (approximately 31 million) and occupies 120 MB of storage. These models were used to assess the trade-off between parameter size and the lightweight model's performance in terms of speed and accuracy.

**TinyViT<sup>[20]</sup>:** TinyViT is a family of compact and efficient vision transformers designed for high performance with minimal computational resources. Pretrained on large-scale datasets, TinyViT employs a fast distillation framework, where a large pre-trained model (teacher) is scaled down into a smaller model (student) while preserving its core capabilities. This student-teacher distillation allows TinyViT to transfer knowledge effectively across various downstream tasks, delivering accurate results with fewer parameters and reduced memory requirements, making it ideal for resource-constrained environments. Similarly, we applied the same fine-tuning approaches to this model. The first approach is designated as TinyViT 20.6M, which contains 20,607,989 parameters and occupies 78.61 MB of storage. The second extended approach is referred to as TinyViT 20.8M, featuring 20,788,277 parameters and requiring 79.30 MB of storage. These experiments were carried out to assess the effect of small changes in

parameter size on performance and efficiency.

### 3-3. Proposed Ensemble Model

This research proposes an ensemble method that integrates CNN-based models (ResNeXt, EfficientNet-B5) with a transformer-based model (TinyViT) to leverage their respective strengths and improve disease classification accuracy as shown in Figure 3. Leveraging multiple models typically boosts accuracy and robustness over individual models.

In this work, each model was underwent trained on the cassava leaf disease dataset and was fine-tuned with different hyperparameter settings. Their classification accuracy was analyzed, and performance-based weights were assigned using the brute force algorithm<sup>[21]</sup>. The brute force approach methodically explores all possible combinations to determine the optimal weights for the ensemble, ensuring accurate and fast results, making it ideal for fewer model combinations.

The final predictions were made using a weighted average of the base models' predictions, with higher weights assigned to more accurate models. This ensemble approach mitigates the weaknesses of individual models by combining their predictions, as illustrated in Figure 3. Each model was trained separately, and their respective weights were used for evaluation. The ensemble prediction utilized various weight combinations to identify the optimal set that maximized accuracy. By incorporating the strengths of each model, this method enhances overall prediction performance rather than disregarding the limitations of any single model.

$$P = \arg \max \sum_{i=1}^n x^i y \quad (1)$$

Here,  $P$  represents the optimal prediction from the ensemble model, where  $x_i$  denotes the weight assigned to each model's output, which is multiplied by the corresponding weight vector  $y$ . The variable  $n$  refers to the total number of models included in the ensemble.

### 3-4. Evaluation Metrics

We evaluate the performance of the models for each

class using various metrics such as precision, recall, and F1-Score, which are calculated using Eq. (2), (3), and (4), respectively. Where True Positives, False Positives, False Negatives, and True Negatives are represented as TP, FP, FN, and TN, respectively. For the overall evaluation of the model's performance on multi-unbalanced classes, we employ the metrics from<sup>[22]</sup>, as described in Eq. (5) rather than the previous average evaluation method<sup>[10][13]</sup>. Where  $N$  refers to the total number of classes, and  $n_i$  represents the fraction of samples of particular class to the total number of samples.  $M_i$  denotes the value of the metric  $M$  for class  $i$ , where the metric  $M$  could be precision, recall, or F1-Score. To compare the performance of the different model variants, we examined the overall recall value. This metric is particularly important in this context, as the cost of missing an infected leaf (FN) is critical, potentially leading to the spread of disease to healthy leaves.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{Overall} = \frac{\sum_{i=1}^N n_i \times M_i}{\sum_{i=1}^N n_i}, \quad (5)$$

We also measured the inference time of various models using two test strategies: centre-crop and multi-crop on the test images. The inference time, representing the duration required for model execution, was recorded in seconds. For the ensemble model, the inference time was measured to be approximately equal to that of the model with the maximum inference time.

**Table 1 Comparison of the two ResNeXt50 model variants, with 23M and 26M parameters, in terms of overall recall (%) using two testing strategies on the multi-class test classification dataset**

Fold	ResNeXt50 23M		ResNeXt50 26M	
	Centre-Crop	Multi-Crop	Centre-Crop	Multi-Crop
1	88.00	<b>88.43</b>	87.34	87.82
2	87.31	87.52	87.61	87.50
3	87.94	88.38	87.73	88.06
4	87.33	87.47	87.73	87.99
5	<b>88.10</b>	88.27	87.78	87.89

**Table 2 Comparison of the two EfficientNet-B5 model variants, with 28M and 31M parameters, in terms of overall recall (%) using two testing strategies on the multi-class test classification dataset**

Fold	EfficientNet-B5 28M		EfficientNet-B5 31M	
	Centre-Crop	Multi-Crop	Centre-Crop	Multi-Crop
1	87.32	88.24	88.48	88.80
2	88.00	88.43	87.96	88.17
3	87.82	88.43	<b>88.57</b>	88.83
4	87.59	88.48	88.36	<b>89.10</b>
5	87.25	87.85	87.96	88.31

**Table 3 Comparison of the two TinyViT model variants, with 20.6M and 20.8M parameters, in terms of overall recall (%) using two testing strategies on the multi-class test classification dataset**

Fold	TinyViT 20.6M		TinyViT 20.8M	
	Centre-Crop	Multi-Crop	Centre-Crop	Multi-Crop
1	89.00	88.99	88.50	88.76
2	<b>89.32</b>	89.32	89.00	88.41
3	89.27	<b>89.51</b>	88.78	89.23
4	88.69	88.69	89.29	89.00
5	88.85	89.15	88.71	88.41

**Table 4 Comparison of various lightweight models and their parameter variants in terms of inference time (seconds) for the two testing strategies on the multi-class test classification dataset**

Model	Inference Time (seconds)	
	Centre-Crop	Multi-Crop
ResNeXt50 23M	1.05734	5.52274
ResNeXt50 26M	1.10840	5.56961
EfficientNet-B5 28M	<b>0.94257</b>	5.20375
EfficientNet-B5 31M	1.35552	<b>4.81913</b>
TinyViT 20.6M	0.95797	5.87840
TinyViT 20.8M	0.99867	5.22033

## 4

## RESULTS AND DISCUSSION

The results of experiments conducted with various lightweight models are detailed in this section. Table 1 compares the centre-crop and multi-crop classification performance on test images using two variants of the ResNeXt50 model with different parameter sizes. The smaller variant, ResNeXt50 23M, achieved the highest overall recall of 88.40% with the multi-crop strategy, outperforming the larger ResNeXt50 26M. However, when compared to ResNeXt, another CNN-based model, EfficientNet-B5 31M, showed an improvement of 0.83%,

achieving an accuracy of 89.10% using the multi-crop strategy, as presented in Table 2. Moreover, despite having more parameters, EfficientNet-B5 required less execution time than ResNeXt, as shown in Table 4. This demonstrates that EfficientNet-B5 not only outperforms ResNeXt in terms of accuracy but also offers greater computational efficiency for multi-class classification tasks.

**Table 5** Display the comparison results of the best-performing lightweight models and our ensemble models across different evaluation metrics for each class and the overall testing dataset

Method	Class	Precision	Recall	F1-score
ResNeXt	CBB	0.6762	0.6544	0.6651
	CGM	0.8112	0.7925	0.8017
	CBSD	0.8624	0.7443	0.799
	CMD	0.9444	0.9738	0.9588
	Healthy	0.7344	0.7287	0.7315
	Overall	0.8822	0.8843	0.8827
EfficientNetB5	CBB	0.7095	0.5853	0.6414
	CGM	0.8271	0.7925	0.8094
	CBSD	0.868	0.7808	0.8221
	CMD	0.9471	0.9802	0.9634
	Healthy	0.7319	0.7461	0.739
	Overall	0.8877	0.891	0.8884
TinyViT	CBB	0.6652	0.6959	0.6802
	CGM	0.7984	0.8134	0.8058
	CBSD	0.8722	0.8105	0.8402
	CMD	0.9581	0.9723	0.9651
	Healthy	0.773	0.7326	0.7522
	Overall	0.8943	0.8951	0.8945
Ensemble	CBB	0.7198	0.6866	0.7028
	CGM	0.8398	0.8134	0.8264
	CBSD	0.8909	0.8014	0.8438
	CMD	0.9556	0.9806	0.9679
	Healthy	0.7713	0.7713	0.7713
	Overall	<b>0.9019</b>	<b>0.9035</b>	<b>0.9023</b>

On the other hand, the vision transformer-based model, TinyViT, outperforms both CNN-based models in terms of inference time and accuracy when using the centre-crop testing strategy, as detailed in Tables 3 and 4. Using the multi-crop strategy, TinyViT achieved 0.41% and 1.08% higher accuracy than ResNeXt and EfficientNet-B5, respectively, with an accuracy of 89.51% using only 20.6M parameters. These results indicate that the lightweight transformer-based model for multi-class classification tasks outperforms the lightweight CNN-based models in terms of efficiency.

Based on the above results, we developed an ensemble model combining the best-performing variants of ResNeXt, EfficientNet-B5, and TinyViT. Interestingly, our multi-crop testing strategy consistently outperformed the centre-crop approach. This can be attributed to the fact that the diseased portion of the image may not always be located in the centre, and relying solely on the centre-crop could result in missing infected regions. To minimize the risk of undetected infections, which could lead to disease spreading to healthy plants, we employed the multi-crop strategy for testing our ensemble model. As shown in Table 5, our ensemble model achieved an F1-score of 0.902 and an overall recall of 90.35%, surpassing the performance of the individual models and setting a new SOTA on the test dataset. The enhanced performance is attributed from the fact that different models tend to make errors on different data samples. One model might excel in learning certain features that others may not. This diversity among the models enhances the robustness of the ensemble, effectively reducing variance in prediction errors and yielding more reliable results.

## 5 CONCLUSION

This research concentrated on building an automated framework to detect and classify cassava leaf diseases at an early stage. The existing methods ensemble either CNN or Transformer models leading to gain in either computational complexity or efficiency. To overcome the challenges of high memory usage and computational demands, we explored various CNN and Transformer-based lightweight models, including ResNeXt and EfficientNet-B5 (CNN), and TinyViT (Transformer). To achieve the best results, we applied centre and multi-crop transformation strategies to the testing dataset and utilized a cross-validation approach to fully utilize the entire dataset for training and testing. We optimized the ensemble method by using a brute force approach to determine the optimal weighted averaging of the models ideal to ensemble few models. Our experimental findings give a novel ensemble model that combines the strengths of transformation strategies with both CNN and Transformer architectures, delivering superior

performance with minimal memory and computational requirements. Although it involves a trade-off in computation time compared to individual models, it remains more efficient than earlier ensemble methods that relied solely on CNN or transformer-based models without considering computational complexity. In future work, we aim to validate this approach on devices like smartphones and agricultural robots for real-time detection of plant diseases in the field.

## REFERENCES

- [1] Huang, K. Y.: Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features. *Computers and Electronics in agriculture* 57(1), 3–11 (2007)
- [2] Taiwo, K. A.: Utilization potentials of cassava in nigeria: the domestic and industrial products. *Food Reviews International* 22(1), 29–42 (2006)
- [3] 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)
- [4] Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D. P.: Deep learning for image-based cassava disease detection. *Frontiers in plant science* 8, 1852 (2017)
- [5] Wang, L., Guo, S., Huang, W., Qiao, Y.: Places205-vggnet models for scene recognition. *arXiv preprint arXiv:1508.01667* (2015)
- [6] Theckedath, D., Sedamkar, R.: Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science* 1(2), 79 (2020)
- [7] Zhang, K., Guo, Y., Wang, X., Yuan, J., Ding, Q.: Multiple feature reweight densenet for image classification. *IEEE access* 7, 9872–9880 (2019)
- [8] Xia, X., Xu, C., Nan, B.: Inception-v3 for flower classification. In: 2017, 2nd ICIVC. pp. 783–787. *IEEE* (2017)
- [9] Mwebaze, E., Gebru, T., Frome, A., Nsumba, S., Tusubira, J.: icassava 2019 fine-grained visual categorization challenge. *arXiv preprint arXiv:1908.02900* (2019)
- [10] Alghamdi, H., Turki, T.: Pdd-net: Plant disease diagnoses using multilevel and multiscale convolutional neural network features. *Agriculture* 13(5), 1072 (2023)
- [11] Zhang, S., Zhang, S., Zhang, C., Wang, X., Shi, Y.: Cucumber leaf disease identification with global pooling dilated convolutional neural network. *Computers and Electronics in Agriculture* 162, 422–430 (2019)
- [12] Lilhore, U. K., Imoize, A. L., Lee, C. C., Simaiya, S., Pani, S. K., Goyal, N., Kumar, A., Li, C. T.: Enhanced convolutional neural network model for cassava leaf disease identification and classification. *Mathematics* 10(4), 580 (2022)
- [13] Kiruthika, V., Shoba, S., Sendil, M., Nagarajan, K., Punetha, D.: Hybrid ensemble-deep transfer model for early cassava leaf disease classification. *Heliyon* (2024)
- [14] Thai, H. T., Tran-Van, N. Y., Le, K. H.: Artificial cognition for early leaf disease detection using vision transformers. In: 2021 International Conference on Advanced Technologies for Communications (ATC). pp. 33–38. *IEEE* (2021)
- [15] Zhuang, L.: Deep-learning-based diagnosis of cassava leaf diseases using vision transformer. In: Proceedings of the 2021 4th Artificial Intelligence and Cloud Computing Conference. pp. 74–79 (2021)
- [16] Chen, J., Zhang, D., Nanekaran, Y. A.: Identifying plant diseases using deep transfer learning and enhanced lightweight network. *Multimedia tools and applications* 79, 31497–31515 (2020)
- [17] Mwebaze, M., Joyce, Julia, D. S.: Cassava leaf disease classification (2020)
- [18] Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
- [19] Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. PMLR (2019)
- [20] Wu, K., Zhang, J., Peng, H., Liu, M., Xiao, B., Fu, J., Yuan, L.: Tinyvit: Fast pretraining distillation for small vision transformers. In: European conference on computer vision. pp. 68–85. Springer (2022)
- [21] Morton, A. B., Mareels, I. M.: An efficient brute-force solution to the network reconfiguration problem. *IEEE Transactions on Power Delivery* 15(3), 996–1000 (2000)
- [22] Grandini, M., Bagli, E., Visani, G.: Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756* (2020)

### ■ 著者



Kaushik Vishwakarma  
Indian Institute of Technology,  
Mandi/  
Emerging Technology and  
Innovation Lab, Yamaha Motor  
Solutions India



Udaybhan Rathore  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India



Rupesh Kumar  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India



Amit Shakya  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India



Chetan Gupta  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India



Shruti Phutke  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India



Lalit Sharma  
Emerging Technology and  
Innovation Lab,  
Yamaha Motor Solutions India