Improved Cassava Plant Disease Classification with Leaf Detection

Ming Xuan Chai*†, Yao Deng Fam*†, Quinito Norman Octaviano*, Chih-Yang Pee*‡,
Lai-Kuan Wong*, Mas Ira Syafila Mohd Hilmi Tan*, John See§
* Faculty of Computing and Informatics, Multimedia University, 63100 Cyberjaya, Malaysia
§ School of Mathematical and Computer Science, Heriot-Watt University (Malaysia), 62200 Putrajaya, Malaysia
† Equal contribution (Emails: mingxuan.chai.cmx@gmail.com and yaodeng07@gmail.com)
‡ Corresponding author (Email: cypee@mmu.edu.my)

***Abstract*—Cassava (Manihot esculenta Crantz), a crucial food source for millions, has experienced significant crop yield losses due to various plant diseases. To address this issue, extensive research has been conducted on the Cassava 2020 dataset released by the Makerere AI Lab. Images in this dataset are captured in complex real-world environments, adding difficulty to the classification process. While several deep learning models have shown promising results, we believe that reducing interference from overlapping leaves and complex backgrounds could further enhance classification performance. To accomplish this, up to five leaves of interest have been identified and annotated in each image. These annotated leaves are then used to train a detection model: Faster R-CNN with ResNet-101, designed to automatically detect leaves of interest in the images. The identified leaves are separated from the background through a masking process, producing masked images (Mi), where i indicates the maximum number of leaves in the masked images, with i ranging from 1 to 5. The masked images M1 to M5 are then trained on various CNN models; i.e. EfficientNetB1, DenseNet121, ResNet50, and Xception, for Cassava disease classification. The results show that classifiers trained with M3 improve the accuracy by 2.13% to 3.06% compared to models trained with original images. The main contribution of this work is a novel technique capable of identifying leaves of interest from complex backgrounds to produce masked images, which improve the accuracy of various CNN-based classification models, and a novel leaf annotation where up to 5 leaves of interest have been annotated for each image in Cassava 2020 dataset. Annotated images can serve as a valuable resource for advanced cassava disease analysis, detection, and classification.**

I. Introduction

Plant diseases pose a serious threat to agricultural productivity, accounting for 20% to 40% of crop losses globally [1]. Therefore, improving plant yields through early detection of plant diseases using advanced artificial intelligence (AI) is crucial to minimize both crop and economic losses [2]. Early plant disease study focus mainly on relatively clean images where traditional machine learning (ML) like K-Means Clustering, Multi-Layer Perceptron (MLP), Support Vector Machine (SVM), k-nearest neighbour(KNN), etc. are effective [3]–[6]. Although these ML algorithms offer promising solutions, effectively detecting diseases in real-world images remain tricky mainly because of mixed healthy and diseased leaves, invisible lesions in the early stages of infection, and complex backgrounds that introduce noise [7].

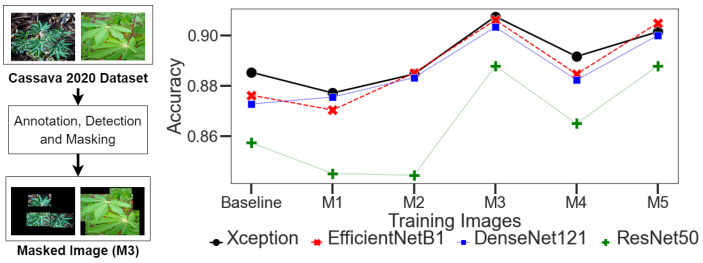


Fig. 1. Accuracy of Cassava disease classification using Xception, EfficientNetB1, DenseNet121, and ResNet50 classifiers trained on original images (Baseline) and masked images (M1-M5). Classifiers trained on M3, M4, and M5 show significant improvements over baseline, while M1 and M2 generally reduce performance, except for DenseNet121. Further details are discussed in Subsection III-B

In order to support advanced machine learning applications in agriculture, particularly for improving the diagnosis and management of cassava diseases, the Artificial Intelligence Lab of Makerere University published the Cassava 2019 dataset in Kaggle [8]. This dataset consists of 22,026 images, where only 5,657 images labeled with diseases. Surya et al. in [9] demonstrated that a pre-trained MobileNetV2 model could fairly classify the diseases but the performance was affected by imbalanced data. Ravi et al. [10] emphasized the disease features on leaves with an attention-based model using EfficientNet and classified the features by implementing meta-classifier (random forest + SVM for prediction and logistic regression for classification). Sambasivam and Geoffrey [11] built a low-cost method to detect cassava infections. The quality of the images was improved by Contrast Limited Adaptive Histogram Equalization (CLAHE), and the images were augmented by applying a combination of SMOTE, focal loss, and class-weight due to the imbalance of the data. In addition, Lilhore et al. [12] enhanced the CNN model by utilizing depth-wise separable convolution, de-correlation stretching with gamma correction, and batch normalization to improve feature count, image color segregation, and overall computational efficiency.

In 2020, the Makerere AI Lab released the Cassava 2020 dataset [13], which serves as an update to the Cassava 2019 dataset. The number of images labeled with diseases in the dataset has grown from 5,657 to 21,397. Utilizing this dataset,

Maryum et al. [14] studied the technique of utilizing U-Net, a deep learning model for leaf segmentation, with EfficientNetB4 (32 layers + 2 fine-tuned layers) for classification with complex backgrounds. The Color Index Vegetation method was applied to extract the vegetation parts from the leaves, providing ground truth for corresponding images. At the same time, Zhang et al. [15] explore a self-supervised learning approach, that allows for classification and training of a reliable model. Their methods utilized the SimCLR model based on ResNet architecture, which maximizes the similarity between the same images with the ResNet18 model to minimize recognition and classification loss during model training. However, self-supervised learning has a critical limitation in training ratios, achieving optimal performance (90%) only with a specific dataset ratio. Concurrently, Zhuang [16] addressed this challenge by utilizing ViT with data augmentation to handle the imbalanced dataset issue and loss function to detect cassava disease on leaves. Singh and the team [17] proposed a similar method using transfer learning techniques with DenseNet169 and the Adam optimizer, resulting in a less time-consuming implementation, though the results are less satisfactory compared to earlier approaches. Soon after, Thai et al. [18] enhanced the base vision transformer with a pruning algorithm to optimize self-attention heads in ViT layers, resulting in reduced model architecture implementation while attaining better results(96.82%) compared to earlier vision transformer approaches.

Inspired by Lilhore et al. [12] and Maryum et al. [14] on how these models reduce background noise in cassava disease classification, we propose a novel framework where an object detection algorithm is used to identify leaves of interest in images so that they can be isolated from images to form masked images. The masked images are then used for training of CNN based cassava disease classifiers like EfficientNetB1, DenseNet121, ResNet50, and Xception. Figure 1 illustrates the forming of masked images from object detection algorithm and performance of CNN based classifiers mentioned above, trained using original images as baseline and masked images (M1-M5). The graph shows compared to baseline, significant improvement in classification accuracy can be achieved when masked images M3 to M5 are used to train the classification models. The main contribution of this article are as follows:

- A novel technique that can accurately identifying leaves of interest from images with complex backgrounds to generate masked images. Classifiers trained using masked images M3 achieve significant improvement in accuracy by 2.13% to 3.06% compared to models trained on original images.
- A new leaf annotation has been created, featuring up to five leaves of interest per image from the Cassava 2020 dataset. This annotation can serve as a valuable resource for advanced disease leaf analysis, detection, and classification.

The remainder of this paper is organized as follows: Section II details the proposed leaf disease classification model, which includes leaf annotations, object detection, masked image generation, and the development of disease classifiers. This is followed by a discussion of experimental results and an ablation study. Finally, the paper concludes in Section IV.

II. Our Approach

In this subsection, the methodology of proposed method is presented. As shown in Figure 2, the process begins with data preparation, where leaf annotation is performed on each image in Cassava 2020 dataset to identify its leaves of interest. The dataset is stratified and split into training and testing sets with a ratio of 80:20. This is following by model building process where leaf detection model is built using the training dataset images together with their leaf annotations. The leaves detection model is then used to identify leaves of interest in each image to produce masked images which are then used to build disease classification system. Finally, the testing images are deployed on the developed leaf detection and disease classification system to evaluate its performance.

A. Cassava 2020 Dataset and leaf Annotation

Cassava 2020 Dataset [13] published by the Makerere Artificial Intelligence Lab in Kaggle, comprises 21,397 labeled images, each with dimensions of 800×600 pixels, a resolution of 96 dpi, and 24-bit RGB colors, stored in JPEG format. The included images consist of healthy (H) cassava plants and four types of diseases: Cassava Bacterial Blight (CBB), Cassava Brown Streak Disease (CBSD), Cassava Green Mottle (CGM), and Cassava Mosaic Disease (CMD). The distribution of the dataset is as shown in TABLE I.

Sample images for each class are shown in Figure 3. As the images are captured in a real-life environment, they are complicated and difficult to classify. Hence, in order to enhance the classification performance, up to five leaves of interest are labelled in each image. The leaf annotation, as shown in Figure 4, are performed by using LabelImg [19] following these guidelines:

1) The first annotated leaf in each image denotes the class to which the image is assigned.
2) Each image must have at least one annotated leaf and no more than five annotated leaves.
3) The order of labeling each leaf is based on its visual prominence.
4) A visually prominent healthy leaf in any disease image should be labeled as healthy (H). Likewise, a visually prominent unhealthy leaf in a healthy image should be labeled as unhealthy (U).

TABLE I
Training and Testing of Cassava 2020 Dataset

Category	Class	Number of Images		
		Training	Testing	Total
0	CBB	870	217	1,087
1	CBSD	1,751	438	2,189
2	CGM	1,909	477	2,386
3	CMD	10,526	2,632	13,158
4	H	2,062	515	2,577
	Total	17,118	4,279	21,397

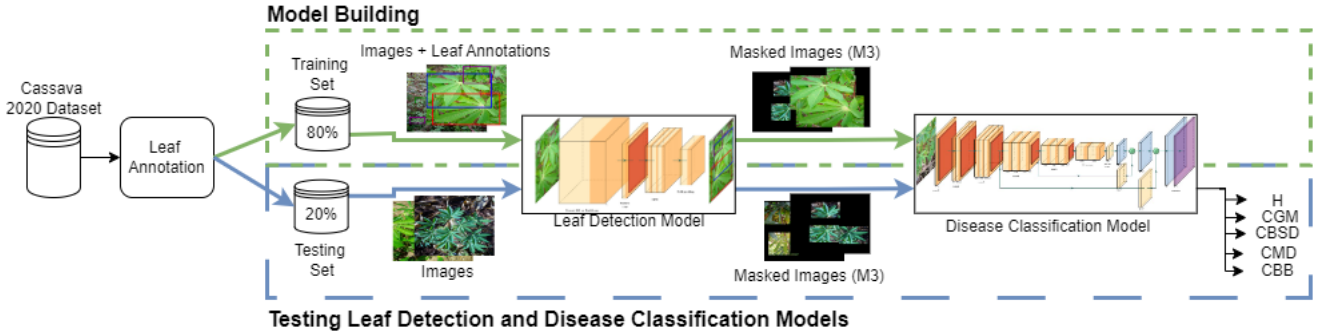


Fig. 2. Framework of Cassava Plant Diseases Classification

Fig. 3. Samples of Cassava Leaves by Classes

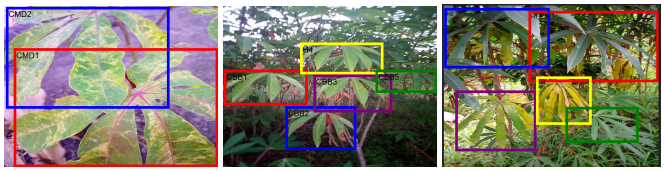


Fig. 4. Sample of Leaf Annotations Where Leaves Sequence are Visualized by Colors of Bounding Boxes: Red, Blue, Purple, Yellow and Green.

An annotated leaf in an image is enclosed by a bounding box and tagged with a label consists of a class: "CBB", "CBSD", "CGM", "CMD", "H", or "U", followed by a number from 1 to 5, indicating the sequence of the leaves. The additional class "U" is used in leaf annotation because prominent unhealthy leaves may appear in otherwise healthy images, as shown in Figure 4(c). Since unhealthy leaves can often appear on healthy plants, they should not be excluded from the image. Most images in the Cassava 2020 dataset focus on a specific leaf (Figure 4(a)) or a small group of leaves (Figures 4(b) and 4(c)). Thus, labeling the five most prominent leaves is typically sufficient to capture the leaves of interest in most Cassava 2020 images.

B. Cassava Leaf Detection and Masked Images

Continued from leaf annotation, three different object detection models; Faster R-CNN with ResNet-101 [20], EfficientDet [21] and Single-Shot Detector (SSD) MobileNetV2 [22]; were evaluated for the leaf detection task to automatically identify leaves of interests from images. Faster R-CNN ResNet101 is known for its object detection prowess. Pre-trained on COCO dataset, the model can generalize to detect various unseen objects in images. SSD EfficientDet and SSD MobileNetV2 offer a compelling balance of detection speed and accuracy, making them the preferred choice. The leaf detection models

TABLE II
MEAN AVERAGE PRECISION OF DETECTION MODELS

	mAP@0.5				
Detection Model	M1	M2	M3	M4	M5
SSD MobileNetV2	0.4817	0.4889	0.5134	0.5288	0.5349
SSD EfficientDet D1	0.5776	0.5982	0.6108	0.6091	0.6005
Faster R-CNN ResNet101	**0.6051**	**0.6136**	**0.6194**	**0.6163**	**0.6050**

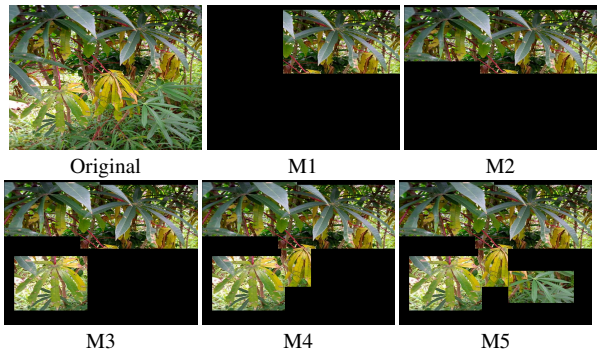


Fig. 5. A Sample of Original and Its Masked Images M1 to M5

were trained on images annotated with different numbers of leaves, ranging from one to five per image, for detecting the same number of leaves. Mean average precision calculated from intersection over union with threshold 0.5 (mAP@0.5) is used to evaluate the performance of leaf detection models. The results are shown in TABLE II where Mi, $i = 1, 2 \cdots, 5$ of columns 2 to 6 denote the scores of mAP@0.5 when up to i leaves of interest are detected by the models in each image.

As shown in the last row of TABLE II, Faster R-CNN ResNet101 outperforms all other models in mAP@0.5 scores in all one to five leaves of interest detection (M1 - M5). In addition, Faster R-CNN ResNet101 score highest in identifying three leaves of interest (M3). Consequently, the leaves detected by Faster R-CNN with ResNet101 are used to create masked images Mi, which are henceforth referred to as masked images containing up to i leaves per image. A sample of original and masked images is shown in Figure 5.

C. Cassava Plant Disease Classification

To evaluate the effectiveness of using masked images generated from leaf detection algorithm, the following CNN models are selected: DenseNet121 [23], EfficientNetB1 [24], ResNet50 [25], and Xception [26]. In this experiment, the CNN models

mentioned above are trained using original Casava training dataset, which serves as the baseline for evaluating the effectiveness of our proposed approach. All models are fine-tuned with pre-trained ImageNet weights [27]. The pre-trained models are connected to a 2D global average pooling layer, followed by the 2 fully connected (FC) layers of 256 and 128 nodes. A dropout rate of 0.4 is applied after each FC layer to allow the models to learn more robust features. A batch normalization layer is inserted after the first dropout layer to speed up the convergence of the model. The final layer of the network utilized Softmax activation function to perform classification of the 5 diseased classes.

III. Empirical Study

A. Experiment Settings

In this experiment, the Cassava 2020 dataset was stratified and split randomly into training and testing sets using 80:20 ratio, with a fixed seed for reproducibility. Input training images were resized to 300×300 pixels. The batch size and epoch were set to 32 and 100, respectively. The initial learning rate was set to 0.001, where it reduced by a factor of 0.1 when there is no improvements in loss for 5 epochs. Early stopping was also utilized to prevent the occurrence of overfitting. The models were trained on a workstation equipped with Intel Core i7-12700K processor running at 3.6 GHz, 64 GB of RAM, and an NVIDIA GeForce RTX 3090 GPU.

B. Classification Performance of the Proposed Model

Our results demonstrate that isolating the leaves from complex backgrounds is effective for improving performance of leaf disease classification. Figure 1 illustrates the performance of Xception, EfficientNetB1, DenseNet121, and ResNet50 classifiers trained on original images (as baseline) and masked images (M1-M5). It can be observed from the results that regardless of the type of CNN models, there is a consistent performance improvement over the respective baseline for models trained with at least 3 leaves of interest. Notably, classifiers trained on M3 images give the best performance. As depicted in row 2 to 5 of TABLE III, the classification models trained on M3 images show significant accuracy improvements of 2.13% to 3.06% over their corresponding baseline models reported in row 6 to 9 of TABLE III respectively. A key highlight is that, across the four CNN models trained on M3, Xception achieved the highest score with 90.64%. Returning to Figure 1, performance of models trained with M5 images is close to M3, with only a slight overall drop in performance. Unexpectedly, there is a consistent performance drop for models trained with M4 images. An investigation into the annotated dataset reveals that some of the smaller leaves, coincidentally those annotated as the fourth leaf, may not accurately reflect the disease class that the image represents. As an example, the fourth labelled leaf (green) shown in Figure 4(b) does not resemble the neigbouring leaves in the same image and it is very likely that the leaf does not belong to the labeled class. This misrepresentation, which is natural in the real-world scenario where not all leaves of a diseased plant show the disease symptoms, is unavoidable and resulted in noise that reduces the models performance. On the other hand, except for DenseNet121, the performance of models trained with M1 and M2 images shows a dip in performance compared to the baseline models. This could be due to the smaller number of leaves of interest in M1 and M2 images, leading to insufficient diversity and information to support the models in learning to discriminate the diseased leaves.

TABLE III
Classification Performance Comparison

Paper	Methods	Train-Test	Accuracy (%)
Proposed Model	**Xception (M3)**	80 - 20	**90.64**
	EfficientNetB1 (M3)	80 - 20	90.61
	DenseNet121 (M3)	80 - 20	90.33
	ResNet50 (M3)	80 - 20	88.79
	Xception (Baseline)	80 - 20	88.51
	EfficientNetB1 (Baseline)	80 - 20	87.60
	DenseNet121 (Baseline)	80 - 20	87.28
	ResNet50 (Baseline)	80 - 20	85.73
Zhang et al. [15]	SimCLR + ResNet18	70 - 30	90.00
Maryum et al. [14]	UNet+ EfficientNetB4	85 - 15	89.90
Singh et al. [17]	DenseNet169	80 - 20	87.86
Zhong et al. [28]	Transformer-Embedded ResNet	80 - 20	**91.12**
Zhuang, L. [16]	Vision Transformer	80 - 20	90.02

Although the baseline models were fine-tuned with the same weights, slight differences in results were observed, especially with ResNet50, which had the lowest accuracy, trailing Xception by 2.78%. This may be due to weaker feature extraction of ResNet50, missing finer details in cassava images with complex backgrounds. In contrast, Xception's architecture, particularly its depthwise separable convolutions, allows it to capture subtle features more effectively. While EfficientNetB1 is highly efficient due to its compound scaling, it may sacrifice sensitivity to intricate details, giving Xception an edge in this task. DenseNet121, though achieving 87.28%, suffers from feature redundancy due to its densely connected architecture, which potentially hinders its ability to focus on new features.

Experiments are compared with existing approaches as detailed in rows 10 to 15 of Table III. Rows 10 to 12 present various CNN approaches trained with the same image size of 224×224 but with different train-test splits. Zhang et al. [15] proposed an ensemble model of SimCLR and ResNet18 with a 70-30 train-test split, achieving an accuracy of 90.00%. Maryum et al. [14] employed U-Net to segment the leaves before classifying the disease with EfficientNetB4, training their model with an 85-15 train-test split dataset, resulting in an accuracy of 89.90%. Furthermore, Singh et al. [17] utilized DenseNet169 to classify cassava leaf diseases, training their model with an 80-20 train-test split dataset, achieving an accuracy of 87.86%. By comparison, the proposed Xception model trained on M3 images demonstrated better performance. Additionally, Table III's rows 13 and 14 present Vision Transformer models trained on higher input image resolution and stratified 5-fold cross-validation, with an 80-20 train-validation split. Zhong et al. [28] achieved 91.12% accuracy using the Transformer-Embedded ResNet with input images sized at 512×512. Additionally, Zhuang [16] trained a ViT model on a dataset with an input image size of 384×384 incorporating image augmentation techniques, achieving 90.02%.

While the proposed Xception method demonstrates comparable performance, its accuracy lags slightly behind Transformer-Embedded ResNet approaches, potentially due to the difference in image resolution.

Overall, fair comparison with state-of-the-art methods is hindered by variations in experimental settings, including image resolution, train-test splits, and evaluation methodology. More crucially, comparison with the baseline demonstrates a considerable improvement in performance, demonstrating the efficacy of our proposed strategy of employing the leaf masks to direct the model's attention.

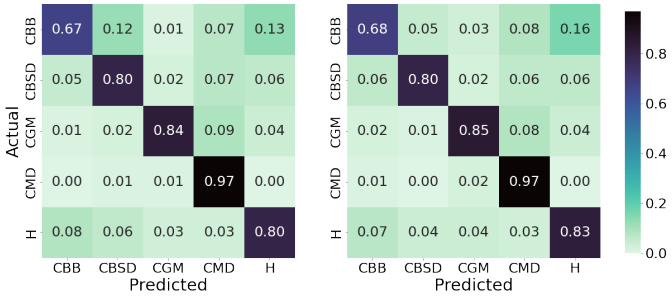
C. Xception and EfficientNetB1 Trained with M3 Images

In this subsection, we analyze the performance of the two best performing models trained on M3 images; Xception and EfficientNetB1, which achieve classification accuracy of 90.64% and 90.61% respectively. Table IV compares the precision, recall, and F1-Score for Xception and EfficientNetB1 models on the five disease classes. It can be observed that the performance of both models are very close, with Exception slightly outperformed EfficientNetB1 on the overall weighted (micro) precision, recall and F1-score. Xception achieves better precision for CBB and H classes, with significant improvements seen in the CBB class, where Xception achieved a precision of 66.06%, which is 3.98% better than EfficientNetB1. On the other hand, EfficientNetB1 yields better precision for CBSD, CGM and CMD. In terms of recall, Xception obtained better scores for CGM and CMD, while EfficientNetB1 scores higher for CBB, CBSD and H. It is interesting that EfficientNetB1 achieves better precision and recall for three of the five classes although it obtained lower scores for the overall metrics. Considering the total number of parameters for EfficientNetB1 (7.1M) is only one-thrid of Xception (22.9M), EfficientNetB1 model can be a better option for resource-constrained deployment.

TABLE IV
XCEPTION AND EFFICIENTNETB1 ON M3 IMAGES

	Xception			EfficientNetB1		
Class	Precision	Recall	F1-Score	Precision	Recall	F1-Score
CBB	0.6606	0.6697	0.6651	0.6208	0.6835	0.6507
CBSD	0.8682	0.7968	0.8310	0.8794	0.7991	0.8373
CGM	0.8244	0.8658	0.8446	0.8269	0.8512	0.8388
CMD	0.9625	0.9761	0.9761	0.9644	0.9677	0.9660
H	0.8422	0.7965	0.8187	0.8356	0.8275	0.8315
Weighted	**0.9076**	**0.9082**	**0.9075**	0.9073	0.9061	0.9064

Figure 6 depicts the confusion matrix for Xception and EfficientNetB1. Results for Exception in Figure 6(a) indicate that the CMD class achieved the highest accuracy at 97.42%, while the CBB class had the lowest accuracy at 67.10%. This discrepancy may be attributed to the imbalanced dataset, with the CMD class being over-represented and the CBB class being underrepresented. Specifically, the CBB class comprises only 1,087 images—approximately 5% of the total dataset—whereas the CMD class includes 13,158 images, representing 60% of the dataset. Notably, the CBB class was misclassified as the Healthy class 13.10% of the time, a relatively high rate compared to other false predictions. This may be due to similarities in patterns between the CBB and Healthy classes, as depicted in Figure 4(b) and Figure 7 that may have contributed to this misclassification. Similarly, the performance of the EfficientNetB1 model, shown in Figure 6(b), reveals a similar pattern, with imbalance data size having an impact on accuracy. However, the misclassification rate for EfficientNetB1 in the CBB and Healthy classes is lower compratively, suggesting that EfficientNetB1 is less sensitive to dataset size variations when compared to Xception model.



(a) Xception (b) EfficientNetB1

Fig. 6. Confusion Matrices for Xception and EfficientNetB1 on M3

(a) H (b) CBB

Fig. 7. Samples of Healthy and CBB Images

D. Xception Classification Performance on Masked Images

Table V compares the results of the Xception model trained on the original images (baseline) and the five set of masked images (M1-M5), categorized by cassava disease classes. The baseline model yielded an overall accuracy score of 88.52%. In contrast, the model trained on M3 images demonstrated a 2.13% improvement, achieving 90.64% accuracy; the model trained on masked images M4, achieved 89.16%, reflecting a 0.65% improvement; and the model trained on masked images M5, improved by 1.61%, reaching 90.12% accuracy. However, models trained with masked images M1 and M2 resulted in slight regressions. This regression could be attributed to the reduced leaf representations, resulting in less diverse and robust leaf features. Notably, the results of all approaches are significantly biased toward CMD, the largest class. Interestingly, isolating leaves from complex backgrounds aids better prediction in minority classes such as CBB and the Healthy class. It can be observed that all models trained with masked images (M1-M5) outperformed the baseline for these two classes. Specifically, CBB achieved 67.10% accuracy on M3 images and the Healthy class achieved 81.01% accuracy with masked images M5. Overall, the proposed method helped to improve the prediction in minority classes within the range of 1.67% to 9.1%. This results reinforce that models trained with at least three leaves of interest (M3-M5) successfully capture sufficient diversity and distinct characteristics of diseased leaves, leading to improved disease classification over the baseline model.

TABLE V
CLASSIFICATION PERFORMANCE OF THE XCEPTION MODEL

	Accuracy(%)					
Class	Baseline	M1	M2	M3	M4	M5
CBB	57.34	63.30	60.55	**67.10**	62.39	61.47
CBSD	80.82	80.14	79.00	80.39	**83.11**	78.09
CGM	78.62	71.07	71.91	**84.41**	80.29	83.86
CMD	**97.45**	96.28	97.07	97.42	96.81	**97.45**
H	72.90	76.12	79.61	80.43	74.76	**81.01**
Overall	88.52	87.71	88.45	**90.64**	89.16	90.12

IV. CONCLUSION

This paper proposes a novel approach to enhance cassava disease classification by implementing the Faster RCNN ResNet101 leaf detection algorithm to identify and extract leaves of interest from images with complex backgrounds. Masked images form by the extracted leaves are subsequently utilized for the training and evaluation of CNN-based classification models, including Xception, EfficientNetB1, DenseNet121, and ResNet50. To train the Faster RCNN ResNet101 detection algorithm, leaf annotations were meticulously conducted on each image from the Cassava 2020 dataset, with up to five of the most prominent leaves being labeled. The experimental results indicate that masked images containing a maximum of three leaves (M3) yield the most significant improvement in disease classification performance, with an increase of approximately 2.13% to 3.06% in accuracy score. Thus, the proposed method effectively extracts local information while mitigating the impact of background complexity, enabling the model to learn more robust class-discriminative features. The leaf annotations can be a valuable resource for advanced cassava disease analysis, detection, and classification.

The proposed approach however faces challenges due to class imbalance in the Cassava 2020 dataset, where larger classes achieve better predictions and smaller classes yield poorer results. Hence, in future work, data augmentation can be applied to the Cassava 2020 images to mitigate the issue of data imbalance, which currently favors majority classes.

ACKNOWLEDGMENT

This work is supported by the Ministry of Higher Education, Malaysia through the Fundamental Research Grant Scheme (FRGS) FRGS/1/2021/ICT06/MMU/03/2.

REFERENCES

[1] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, "Global food demand and the sustainable intensification of agriculture," Proceedings of the national academy of sciences, vol. 108, no. 50, pp. 20260–20264, 2011.

[2] H. Mishra and D. Mishra, "Artificial intelligence and machine learning in agriculture: Transforming farming systems," Research Trends in Agriculture Science, vol. 1, pp. 1–16, 2023.

[3] F. Faithpraise, P. Birch, R. Young, J. Obu, B. Faithpraise, and C. Chatwin, "Automatic plant pest detection and recognition using k-means clustering algorithm and correspondence filters," 2013.

[4] M. Ebrahimi, M. H. Khoshtaghaza, S. Minaei, and B. Jamshidi, "Vision-based pest detection based on svm classification method," Computers and Electronics in Agriculture, vol. 137, pp. 52–58, 2017.

[5] A. Ojha and V. Kumar, "Image classification of ornamental plants leaf using machine learning algorithms," in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 834–840, 2022.

[6] S. M. Javidan, A. Banakar, K. A. Vakilian, and Y. Ampatzidis, "Diagnosis of grape leaf diseases using automatic k-means clustering and machine learning," Smart Agricultural Technology, vol. 3, p. 100081, 2023.

[7] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning—a review," IEEE Access, vol. 9, pp. 56683–56698, 2021.

[8] E. Mwebaze, T. Gebru, A. Frome, S. Nsumba, and J. Tusubira, "icassava 2019 fine-grained visual categorization challenge," arXiv preprint arXiv:1908.02900, 2019.

[9] R. Surya and E. Gautama, "Cassava leaf disease detection using convolutional neural networks," in 2020 6th international conference on science in information technology (ICSITech), pp. 97–102, IEEE, 2020.

[10] V. Ravi, V. Acharya, and T. Pham, "Attention deep learning-based large-scale learning classifier for cassava leaf disease classification," Expert Systems, vol. 39, 11 2021.

[11] G. Sambasivam and G. D. Opiyo, "A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks," Egyptian Informatics Journal, vol. 22, no. 1, pp. 27–34, 2021.

[12] U. K. Lilhore, A. L. Imoize, C.-C. Lee, S. Simaiya, S. K. Pani, N. Goyal, A. Kumar, and C.-T. Li, "Enhanced convolutional neural network model for cassava leaf disease identification and classification," Mathematics, vol. 10, no. 4, p. 580, 2022.

[13] E. Mwebaze, J. Mostipak, Joyce, J. Elliott, and S. Dane, "Cassava leaf disease classification," 2020.

[14] A. Maryum, M. U. Akram, and A. A. Salam, "Cassava leaf disease classification using deep neural networks," in 2021 IEEE 18th international conference on smart communities: improving quality of life using ICT, IoT and AI (HONET), pp. 32–37, IEEE, 2021.

[15] H. Zhang, Y. Xu, and J. Sun, "Detection of cassava leaf diseases using self-supervised learning," in 2021 2nd International Conference on Computer Science and Management Technology (ICCSMT), pp. 120–123, IEEE, 2021.

[16] L. Zhuang, "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.

[17] R. Singh, A. Sharma, N. Sharma, and R. Gupta, "Automatic detection of cassava leaf disease using transfer learning model," in 2022 6th International Conference on Electronics, Communication and Aerospace Technology, pp. 1135–1142, IEEE, 2022.

[18] H.-T. Thai, K.-H. Le, and N. L.-T. Nguyen, "Formerleaf: An efficient vision transformer for cassava leaf disease detection," Computers and Electronics in Agriculture, vol. 204, p. 107518, 2023.

[19] Tzutalin, "Labelimg." Free Software: MIT License, 2015.

[20] B. Cheng, Y. Wei, H. Shi, R. Feris, J. Xiong, and T. Huang, "Revisiting rcnn: On awakening the classification power of faster rcnn," in Proceedings of the European conference on computer vision (ECCV), pp. 453–468, 2018.

[21] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10781–10790, 2020.

[22] Y.-C. Chiu, C.-Y. Tsai, M.-D. Ruan, G.-Y. Shen, and T.-T. Lee, "Mobilenet-ssdv2: An improved object detection model for embedded systems," in 2020 International conference on system science and engineering (ICSSE), pp. 1–5, IEEE, 2020.

[23] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.

[24] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in International conference on machine learning, pp. 6105–6114, PMLR, 2019.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[26] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.

[28] Y. Zhong, B. Huang, and C. Tang, "Classification of cassava leaf disease based on a non-balanced dataset using transformer-embedded resnet," Agriculture, vol. 12, p. 1360, 09 2022.