Research Article

Received: 19 March 2024 Revised: 9 September 2024 Published online in Wiley Online Library:

(wileyonlinelibrary.com) DOI 10.1002/ps.8456

Cassava disease detection using a lightweight modified soft attention network

Arailym Dosset,[a] L. Minh Dang,[b,c,d] Faisal Alharbi,[e] Shabana Habib,[f] Nur Alam,[a] Han Yong Park[g] and Hyeonjoon Moon[a*]

Abstract

BACKGROUND: Cassava is a high-carbohydrate crop that is at risk of viral infections. The production rate and quality of cassava crops are affected by several diseases. However, the manual identification of diseases is challenging and requires considerable time because of the lack of field professionals and the limited availability of clear and distinct information. Consequently, the agricultural management system is seeking an efficient and lightweight method that can be deployable to edged devices for detecting diseases at an early stage. To address these issues and accurately categorize different diseases, a very effective and lightweight framework called CDDNet has been introduced. We used MobileNetV3Small framework as a backbone feature for extracting optimized, discriminating, and distinct features. These features are empirically validated at the early intermediate stage. Additionally, we modified the soft attention module to effectively prioritize the diseased regions and enhance significant cassava plant disease-related features for efficient cassava disease detection.

RESULTS: Our proposed method achieved accuracies of 98.95%, 97.03%, and 98.25% on Cassava Image Dataset, Cassava Plant Disease Merged (2019–2020) Dataset, and the newly created Cassava Plant Composite Dataset, respectively. Furthermore, the proposed technique outperforms previous state-of-the-art methods in terms of accuracy, parameter count, and frames per second values, ultimately making the proposed CDDNet the best one for real-time processing.

CONCLUSION: Our findings underscore the importance of a lightweight and efficient technique for cassava disease detection and classification in a real-time environment. Furthermore, we highlight the impact of modified soft attention on model performance.

Keywords: cassava disease detection; crop; deep learning; pest control; modified attention; lightweight network

1 INTRODUCTION

Cassava is the primary source of food among different groups of root tubers, especially those grown in tropical regions like South America, East and West Africa, and South Asia.[1] It is considered the fourth significant source of carbohydrates after rice, wheat, and maize.[2] According to the Food and Agriculture Organization of the United Nations (FAO), Nigeria is the world's largest cassava producer, contributing ∼60 million metric tonnes to global production.[3] Additionally, the FAO estimates that global cassava production reached a total of 302.66 million metric tons. The cultivation of this crop is progressively growing in Africa due to its diverse forms of utilization for human consumption and ability to effectively adjust to environmental challenges where other crops often struggle to thrive.[4] In addition, ∼70 million people in this region rely on cassava as their main source of food, providing over 500 kcal per person every day.[5] It is widely utilized as animal feed, used as raw material for related industries, and is consumed as food due to its crucial vitamins, including vitamins B and C, and other essential minerals.[6]

However, various kinds of diseases and pests[7] affect cassava growth and development, cause significant economic losses for many smallholder farmers, and reduce crop yields. There are

* Correspondence to: H Moon, Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea. E-mail: hmoon@sejong.ac.kr

a Department of Computer Science and Engineering, Sejong University, Seoul, Republic of Korea

b Institute of Research and Development, Duy Tan University, Da Nang, Vietnam

c Faculty of Information Technology, Duy Tan University, Da Nang, Vietnam

d Department of Information and Communication Engineering and Convergence Engineering for Intelligent Drone, Sejong University, Seoul, Republic of Korea

e Quantum Technologies and Advanced Computing Institute, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia

f Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

g Department of Bioresource Engineering, Sejong University, Seoul, Republic of Korea

diverse types of cassava diseases that have been reported in multiple places around the world.[8] The pathogens *Xanthomonas axonopodis pv. Manihotis* and African Cassava Mosaic Virus (ACMV) are ranked as the sixth and seventh most significant pathogenic microorganisms in worldwide significant diseases, respectively.[9] They are responsible for causing cassava bacterial blight and cassava mosaic disease (CMD).[9] CMD is the most widespread disease and is considered a significant challenge to cassava cultivation, with a total crop loss of US$1.9–2.7 billion annually.[10] Currently, CMD in African and Asian regions and cassava brown streak disease (CBSD) in Africa are considered the most serious global threats to cassava production.[11] To overcome this production loss, early detection of cassava diseases is important for agricultural industries to overcome the disease's propagation and improve crop yields. To address this issue, several common methods have been proposed. Due to the advancement of technology, Deep Learning (DL) algorithms are becoming more popular due to their excellent performance, and the application of this technique has become a hot topic among researchers within the domain of cassava disease detection. For example, Ramcharan *et al.*[12] applied transfer learning to the InceptionV3 to detect cassava diseases and Sangbamrung *et al.*[13] presented a multistage method for CBSD disease identification. First, a convolutional neural network (CNN) was applied to classify unhealthy and healthy cassava images, then Faster Region-Convolutional Neural Network (Faster R-CNN) was utilized to detect CBSD disease in unhealthy classes. However, it is challenging to check the robustness of this method due to its inability to classify the other types of diseases except CBSD. Sambasivam and Opiyo[14] trained the CNN model from scratch for the classification of the cassava dataset. This study used the synthetic minority oversampling technique (SMOTE), class weight, and focus reduction techniques to address the issue of class imbalance in the dataset, but applying this proposed method to the complicated dataset is challenging. Abayomi-Alli *et al.*[15] expanded the training data by generating synthetic images with modified color histograms. They presented four different approaches to reduce image quality.

Over the past few years, numerous studies have focused on detecting and categorizing cassava diseases. Table 1 discusses and reviews the recent studies in this field. Riaz *et al.*[16] proposed the EfficientNetB3 model to categorize the Kaggle competition 2020 cassava dataset. They applied the data augmentation method to obtain equal samples for each class and achieved overall accuracies of 83.03% without augmentation and 84.48% after augmentation for the test dataset. Oyewola *et al.*[17] introduced a novel deep residual convolution neural network (DRNN) to recognize CMD in cassava leaves and classification of diseases. Specific attention is given to block processing techniques, which may involve dividing the image data into distinct blocks for balancing the dataset. Lastly, DRNN was compared with the plain CNN model, and recognition was successfully achieved with 96.75% accuracy. Hassan and Maji[18] introduced the inception layer and residual connection-based CNN to identify diseases in cassava leaf images. Moreover, they applied depthwise separable convolution to decrease the parameter count. To check the performance of the model they utilized three plant datasets and achieved the highest accuracy of 76.59%. Liu and Liang[19] concentrated on identifying cassava diseases using a multiscale fusion model that integrates the EfficientNet backbone model and an attention module. The proposed method was performed with 88.1% accuracy. Ravi *et al.*[20] integrated attention-based approach into of a pre-trained CNN-based EfficientNet models to accurately detect and locate tiny diseased cassava leaf regions. The proposed method achieved between 65% and 88% for recall, precision, and F1 score macros. Maryum *et al.*[21] suggested the EfficientNetB4 model to classify the Kaggle dataset, which contains 21 397 images of cassava leaves. The U-Net semantic segmentation approach was applied to obtain only leaves from the images. Finally, classification was performed with 81.48% and 89.09% for original and masked images, respectively. The authors presented a novel federated learning CNN and utilized a dataset of cassava leaf images, which were classified into four severity categories. The final accuracy of the federated learning CNN model was reported to be 95%. However, the model struggled to distinguish between mild and moderate severity levels because of tiny visual differences. Mehta *et al.*[22] and Lilhore *et al.*[23] introduced the enhanced CNN method to classify the Kaggle cassava dataset. The dataset was normalized and categorized into two main classes: healthy and unhealthy, and the SMOTE method was used for resampling purposes. The enhanced CNN (ECNN) model outperformed the existing traditional CNN model and achieved 99.3% accuracy for the augmented dataset.

However, despite the compelling results of various DL approaches, the reliability of cassava disease detection needs improvement. In addition, these models are computationally intensive and require powerful Graphics Processing Unit (GPU) and Tensor Processing Unit (TPU) for efficient processing. Moreover, there is a need for a system that should be not expensive and covers a large area. Our study proposed an efficient lightweight CNN model to address these problems. The significant contributions of our work are presented as follows:

- The detection of numerous diseases regarding crops is a challenging task due to their diverse sizes, forms, colors, lighting conditions, and locations. These factors also lead to inaccurate detection outcomes of Computer Vision Assisted (CVA) solutions. As a result of this research, CDDNet has developed a lightweight system capable of detecting agricultural diseases and accurately recognizing them.
- To prioritize attention on intermediate features, we made modifications to the Soft Attention (SA) module to specifically target the complex diseased regions. The gradual refinement of the suggested model enhances the detection of affected regions with greater accuracy, resulting in a superior prediction of the baseline.
- The CDDNet can effectively analyze and categorize deep discriminative data extracted from an infected cassava leaf. An extensive empirical investigation was carried out on the suggested model to examine the duration of training, dimensions, and computational requirements, demonstrating its compatibility with devices that have limited resources.
- To determine the efficiency of our proposed network, we employed three complex datasets: Cassava Image Dataset, the Cassava Plant Disease Merged (2019–2020) Dataset, and a new Cassava Leaf Disease Combined Dataset. The experimental findings demonstrated the superior performance of our proposed CDDNet compared to the state-of-the-art (SOTA) methods.

This article is presented in the following way. Section 2 discusses the detailed information about the proposed dataset and the architecture of our proposed method. Subsequently, Section 3 delves into the explanation of experimental findings and qualitative results, and highlights the time-complexity analysis of the

Table 1. Summary of related works[12–23] in this study

Study	Overview	Method	Dataset availability	Accuracy
Ramcharan *et al.*[12]	The authors proposed the InceptionV3 CNN model to classify the cassava dataset	InceptionV3	✓	93%
Sangbamrung *et al.*[13]	The authors first proposed the CNN model to classify healthy and unhealthy cassava leaves, then used faster R-CNN to detect CBSD	CNN and Faster R-CNN	×	96%
Sambasivam *et al.*[14]	The authors proposed to train the CNN model from scratch and used class weight, SMOTE, and focal loss techniques	CNN	✓	93%
Aboyami-Alli *et al.*[15]	The authors in this study used MobilenetV2 on an improved dataset	MobileNetV2	✓	99.7%
Riaz *et al.*[16]	The authors proposed the EfficientNetB3 model and used data augmentation techniques	EfficientNetB3	✓	83.03%
Oyewola et al[17]	The authors proposed the DRNN model and showed that DL with residual connection outperformed plain CNN	DRNN	✓	96.75%
Hassan & Maji[18]	The authors proposed the CNN model based on the inception layer and residual connection	CNN based on Inception and residual connection	✓	76.59%
Liu & Liang[19]	The authors proposed an attention-driven network for a multiscale fusion	Attention-based EfficientNet	✓	88.1%
Ravi *et al.*[20]	The authors focused attention on the EfficientNet model	Attention-based EfficientNet	✓	65–88% marc and f1-score
Maryum *et al.*[21]	The authors proposed the EfficientNetB4, then the U-Net segmentation approach was used to obtain only leaves from the images	EfficientNetB4	✓	89.09%
Mehta *et al.*[22]	The authors proposed federated learning CNN	Federated learning CNN	×	95%
Lilhore *et al.*[23]	The authors proposed the enhanced CNN model	ECNN	✓	99.3%

The ✓ indicates the dataset is publicly available while × represents datasets with restricted access.

Table 2. The statistical overview of the benchmark datasets, including Cassava Image, Cassava Plant Disease Merged (2019–2020), and the newly established Cassava Leaf Disease Combined datasets

Dataset	Training	Testing	Total
Cassava Image Dataset	7197	1799	8996
Cassava Plant Disease Merged (2019–2020) Dataset	21 642	5411	27 050
Cassava Leaf Disease Combined Dataset	47 046	11 761	58 807

proposed model in comparison to other baseline models. Finally, in Section 4 we conclude the article with some remarks and suggestions for future directions.

2 MATERIALS AND METHODS

2.1 Materials

We employed two datasets, Cassava Image Dataset[12] and the Cassava Plant Disease Merged (2019–2020) Dataset, to make the proposed CDDNet model perform effectively. Furthermore, to make the proposed model considerably more robust, the Cassava Leaf Disease Combined Dataset was developed by merging the two previous datasets. Comprehensive statistical information about the used datasets is presented in Table 2 and the sample images are given in Fig. 1.

2.1.1 *Cassava Image Dataset*

Leaf images of cassava were captured using a 20.2 MP Sony Cybershot camera for experiments belonging to the International Institute of Tropical Agriculture (IITA) in Africa. A total of 11 670 images were captured within 4 weeks. Each disease or kind of damage caused by pests exhibited distinct characteristics, with minimal diversity in symptom manifestation among different varieties compared to the extreme disparities between diseases. Subsequently, the images were examined to identify cases of co-infections, reducing the inclusion of images depicting several diseases. The dataset, referred to as the 'original cassava dataset,' consisted of a total of 2756 images. The original dataset consisted of six class labels, which were assigned by manual allocation after in-field diagnoses carried out by cassava disease specialists from the IITA. The original dataset consisted of the

Figure 1. Images of samples from Cassava Image Dataset,[12] the Cassava Plant Disease Merged (2019–2020)[30] Dataset and the Cassava Leaf Disease Combined Dataset utilized in this study. CBLS, cassava brown leaf spot; CBSD, cassava brown streak disease; CGM, cassava green mite; CHL, cassava healthy leaf; CMD, cassava mosaic disease; CRM, cassava red mite.

following disease classes and their respective number of images: CBSD 398 images, CMD 388 images, and cassava brown leaf spot (CBLS) 386 images. Additionally, there were two classes for mite damage: cassava green mite (CGM) 309 images and cassava red mite (CRM) 415 images. Lastly, there was a class for cassava healthy leaf (CHL) with 353 images. We enhanced the dataset through the application of image augmentation techniques, including rotations at angles of 90°, 180°, and 270°, as well as the implementation of color saturation adjustments. The enhancements aimed to bolster the model's robustness and enhance its potential to be applied across various scenarios. The dataset comprised 8996 photos after augmentation and color saturation. We further partitioned it into training and testing sets, with 80% allocated to training and 20% to testing.

2.1.2 *Cassava Plant Disease Merged (2019–2020) Dataset*

This dataset contains photos of the cassava plant, the most significant supplier of carbohydrates in Africa. Farmers place great emphasis on this crop because of its resilience under adverse circumstances. The dataset consists of five categories of cassava diseases. The dataset initially included 1553 image of CRM, 3632 images of CBSD, 3159 images of CGM, 15816 images of CMD, and 2893 healthy cassava images. Due to the dataset's intrinsic imbalance, our model initially showed reduced accuracy. We used several image augmentation approaches to create a more balanced dataset and improve our model's capacity to generalize under various scenarios. After augmentation, the dataset contained 27 053 images. The dataset, consisting of 27 053 images, was divided into distinct training and evaluation splits. More precisely, around 21 642 images (80% of the total images) were designated for training. The testing set was assigned the remaining 20%, around 5411 images.

2.1.3 *Cassava Leaf Disease Combined Dataset*

We combined Cassava Image Dataset and Cassava Plant Disease Merged (2019–2020) Dataset to create a new and more intricate dataset to examine the robustness of our method. The Cassava Leaf Disease Combined dataset encompassed a total number of 58 807 images. The extensive variety of cassava leaf species in the new dataset contributes to its uniqueness and complexity, which consequently renders the model training procedure highly rigorous. The generalizability and dependability of the resulting model for real-time cassava disease recognition scenarios were improved.

2.2 Methods

The suggested network comprises three primary phases. First, the cassava plant leaf images that have been obtained were subjected to preprocessing. Subsequently, the various categories' images were directed into a proficient proposed framework that accurately detected and categorized the cassava plant into its corresponding classes. Finally, the model made its decision according to the expected class for the provided image. In terms of practical implementation, when the system predicts that a plant is affected by a disease, it generates an alert to the nearby plant disease management department, prompting it to take immediate action. The proposed framework's graphic representation is presented in Fig. 2.

2.2.1 *Backbone features extractor*

Over the past few years, CNN networks for mobile devices have undergone significant development, starting from 2017 and continuing onwards. During this period, three distinct MobileNet architectures were consistently improved. During the development of MobileNetV1, the typical VGG network was consulted and depthwise discrete convolutions were incorporated. By using this information, MobileNetV2[24] was launched 1 year later, featuring an inverted residual and a linear bottleneck. MobileNetV3 underwent improvements in 2019 by eliminating expensive layers and replacing the rectified linear unit (ReLU) function with the h-swish non-linearity function. These enhancements were achieved through neural architecture search and AdaptNet network searching for architectural optimization. By adopting these modifications, MobileNetV3 became more efficient and achieved higher relative accuracy.

2.2.2 *Depthwise separable convolution*

The MobileNetV3 includes a unique computing technique depthwise separable convolution. This convolution method differs from standard convolutions as it performs convolutional computations for every single layer in two different phases. In the initial stage of depthwise convolution, each input channel is subjected to a separate convolutional filter. Subsequently, in depthwise convolution, the generated output channels are integrated into pointwise channels through convolutional processing. Even though there are some losses in accuracy, the depthwise separable convolution makes computing more efficient by reducing the time needed for processing. The utilization of depthwise separable convolutional parameters significantly contributes to the improvement of model efficiency, such as MobileNetV1-V3 and related variations.

2.2.3 *Linear bottlenecks*

MobileNetV3 employs linear bottlenecks (LBs) to reduce the dimensionality of the input data. This method facilitates the retrieval of features from a space of high dimensionality while ensuring minimal loss of information. The term 'linear bottleneck' denotes a Convolutional Layer (CL) that employs a linear activation function in conjunction with a filtering layer. MobileNetV3 incorporates LB layers within convolutional blocks as an alternative to conventional ReLU function transformations due to their inclusion of nonlinearity and the possibility of information loss.

2.2.4 *Inverted residual*

The bottleneck layers operate as an improved and more efficient alternative to the ReLU layers for retrieving extensive information. Additionally, an expansion layer is present as part of the bottleneck block. The mobile network efficiently transmits gradients between layers by establishing direct connections between

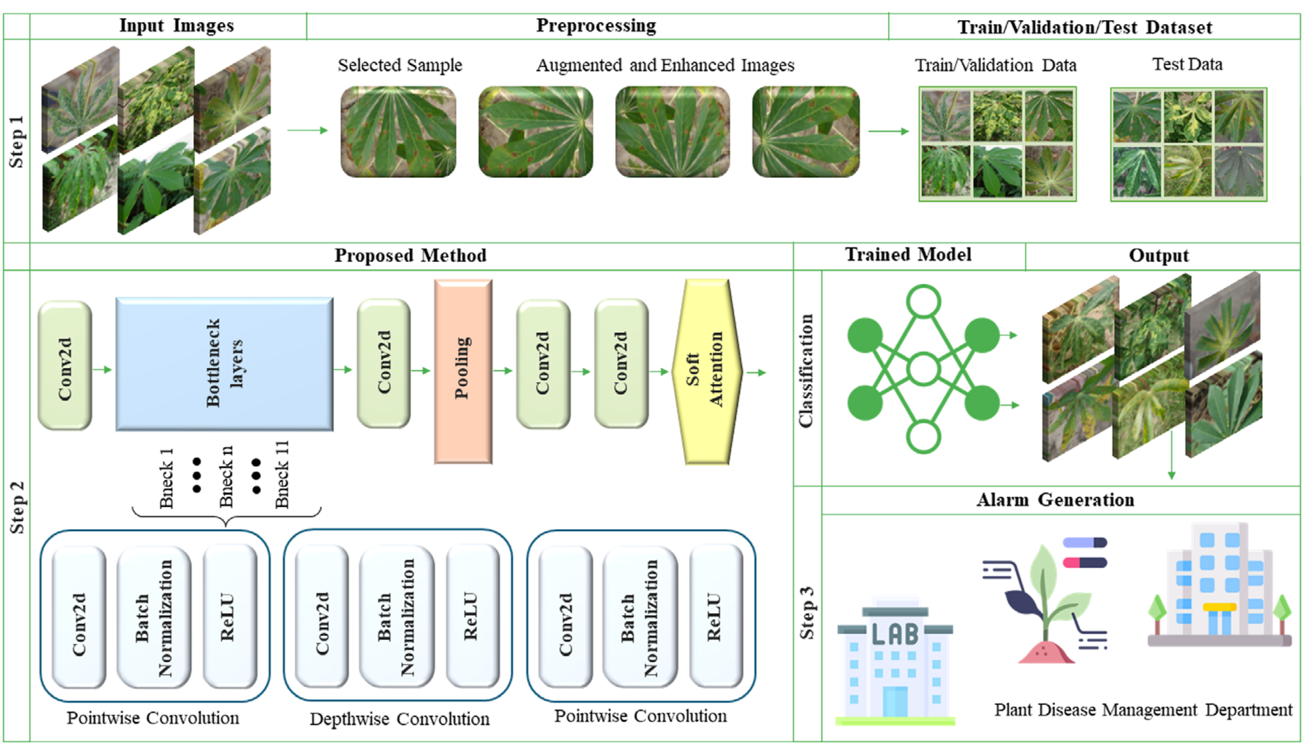


Figure 2. The main framework of the proposed CDDNet approach.

bottlenecks, thus preventing the loss or explosion of gradients across layers. Inverted residual blocks exhibit comparable performance to residual blocks while achieving a notable reduction in memory consumption.

2.2.5 Neural architecture search

Reinforcement learning and Recurrent Neural Network (RNN) are employed to determine the most efficient structure for MobileNetV3 on a limited hardware platform. For instance, the enhanced layer of MobileNetV3 has been restructured using the initial concept of MobileNetV2. A new activation function called Swish has emerged as a replacement for the ReLU function, offering improved accuracy levels. This function is defined as follows:

$$\mathrm{Swish}(a) = a \times \sigma(a) \tag{1}$$

Here, a represents the input to the activation function, and $\sigma(a)$ is the sigmoid function applied to a. The delicate attributes of the sigmoid function utilized in Swish may demand a significant amount of processing resources on portable devices. MobileNetV3 is a proposed approach to address this problem. To accomplish this, the sigmoid function in Swish is approximated using the ReLU6 function. The term used to describe this is h-Swish, which is precisely described as:

$$\mathrm{h{-}swish}(a) = a\,\frac{\min(\max(a+3,0),6)}{6} \tag{2}$$

The expression $a + 3$ moves the input a by 3 units to the right. The max function limits the output to 0 if the value of $a + 3$ is negative. The main function restricts the previous result to a maximum value of 6. Subsequently, the result is divided by 6 to standardize it within the range of 0–1.

2.2.6 Modified soft attention

To tackle the issue of targeting specific locations affected by cassava leaf disease, we integrated a soft attention approach into the prevailing MobileNetV3small design. This method is based on the insight that intermediate traits necessitate a higher level of attention. To tackle this issue, we utilized modified soft attention (MSA), which allows us to selectively highlight important sections of the image while diminishing less significant areas. Traditional CNNs distributed computing resources evenly across the entire image without considering the variable importance of different regions. Our MSA technique incorporates an attention gate, deliberately selected three-dimensional (3D) CLs, as well as maximum pooling operations. The adaptive attention gate approaches fine-tune feature values, empowering the model to zero in on essential patterns within cassava categories, especially in differentiating between affected and healthy plants. This strategy efficiently reallocates computing resources to targeted locations, improving the model's capacity to identify essential features. This technique improves processing accuracy and speed by reducing needless calculations and emphasizing the unique features of input photographs. Integrating two more 3D layers into the soft attention module improves the model's capacity to comprehend intricate relationships in many dimensions. Additionally, using maximum pooling with a 2×2 size and ‘same’ padding on the combined features helps in effective downsampling while preserving spatial characteristics. This method is essential for separating significant characteristics and maintaining the spatial coherence of the pictures. Using ‘same’ padding ensures uniformity in dimensions, aligning well with our systematic feature engineering methodology.

$$g(c\alpha) = \rho m\left(\left(\sum_{l=1}^{L} softmax(\mathbf{W}l \times s)\right)\right) \tag{3}$$

The tensor $m\ \Upsilon^{h \times w \times d}$ is used as the commencing input for a 3D CL. This layer is identified by its weight scheme φ_ν, extending over the dimensions $^{h \times w \times d \times \tau}$, where τ signifies the 3D filter weights. After the convolution process, the resulting output is modified by a softmax function, creating 16 individual attention maps represented by τ. These maps are collectively merged into a single, aggregate attention map, which acts as a weighting function, represented by the symbol β. The tensor m is extensively updated by α and adjusted by a trainable variable ρ. The precisely tuned characteristics g($c\alpha$) are added to the basic features tensor m into a residual branch. The variable ρ is set to 0.0001 during training so that the model updates its weights gradually, which leads to a more stable convergence toward the minimum loss function. This allows the network to gradually adjust and control the required amount of attention for maximum efficiency.

3 RESULTS AND DISCUSSION

This section starts by discussing the experimental configuration and evaluation parameters. It then proceeds with assessing the results obtained from the experiment. All the models involving the proposed method are trained by employing 30 epochs in total and a low learning rate. Each model is retrained based on the results using its originally specified input size. The training is conducted using a batch size of 16. The Adam optimizing algorithm is employed with a learning rate set as 1e-4 and a momentum value of 0.9. The experiments are carried out using an NVIDIA RTX 3090 GPU with 8 GB of RAM. The Keras DL framework is used with TensorFlow as the backend. The proposed model's efficacy is assessed using several assessment metrics, such as precision, recall, accuracy, and F1 score, as shown in the following equations.

3.1 Evaluation metrics

The performance of the model is evaluated based on key components of the diffusion matrix, which include True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Classification accuracy describes the ratio of correct predictions among all predictions:

$$\mathrm{accuracy} = \frac{TP + FN}{TP + TN + FP + FN} \tag{4}$$

Precision is a metric that assesses the accuracy of classifying cassava disease in a dataset by describing the proportion of instances that were accurately labeled as disease. The anticipated positives are images categorized as illnesses, including both TP and FP. Images showing a disease scenario are classified as TP.

$$\mathrm{precision} = \frac{TP}{TP + FP} \tag{5}$$

Recall is a statistical indicator that indicates the proportion of instances in a dataset that were correctly identified as having cassava plant disease by the model. The algorithm effectively forecasts TP for plant disease images.

Table 3. Comparison of input size and number of parameters used for training of the proposed method with SOTA methods

Model	Input size	Batch size	Parameters
VGG19[25]	224 × 224	16	143.67
VGG16[25]	224 × 224	16	138.36
EfficientNetB0[26]	224 × 224	16	5.33
MobileNetV2[27]	224 × 224	16	3.54
ResNet50[28]	224 × 224	16	25.64
NasNetMobile[29]	224 × 224	16	5.33
Proposed method	224 × 224	16	1.67

$$\text{recall} = \frac{TP}{TP + FN} \tag{6}$$

The F1 score evaluates the proportional average of accuracy and recall.

$$F1\text{-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{7}$$

3.2 Comparison with SOTA methods

This study presents the comparison of our proposed model with other pretrained CNN architecture for identifying diseases in

Table 4. Performance comparative analysis of the suggested approach with SOTA methods

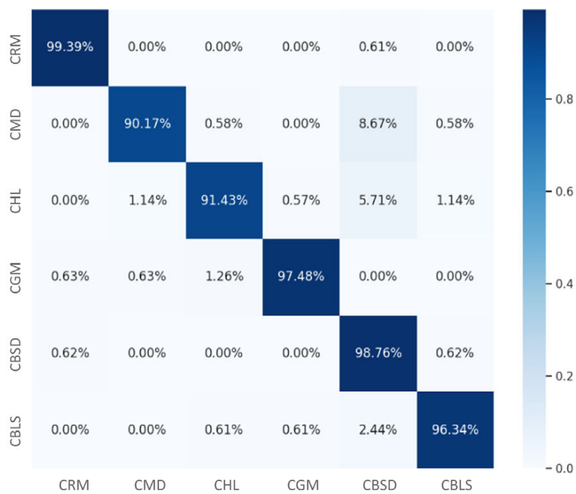
		Cassava Image Dataset				Cassava Plant Disease Merged (2019–2020) Dataset				Cassava Leaf Disease Combined Dataset			
Model	Class	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score	Accuracy
VGG19[25]	CRM	0.95	0.95	0.95	96.59	0.95	0.94	0.94	94.08	0.95	0.94	0.94	94.83
	CMD	0.97	0.88	0.88		0.87	0.95	0.91		0.88	0.95	0.91	
	CHL	1	0.96	0.96		0.94	0.99	0.96		0.94	1	0.97	
	CGM	0.89	0.99	0.99		0.96	0.89	0.92		0.98	0.89	0.93	
	CBSD	0.97	0.99	0.99		0.95	0.95	0.95		0.96	0.97	0.97	
	CBLS	0.99	0.97	0.97		–	–	–		0.97	0.96	0.97	
VGG16[25]	CRM	0.85	0.94	0.89	94.58	0.93	0.85	0.89	92.72	0.94	0.85	0.89	93.45
	CMD	0.91	0.93	0.92		0.91	0.91	0.91		0.93	0.9	0.92	
	CHL	0.95	0.98	0.96		0.95	0.95	0.94		0.95	0.95	0.95	
	CGM	0.99	0.91	0.95		0.97	0.97	0.94		0.92	0.98	0.95	
	CBSD	1	0.92	0.96		0.97	0.97	0.93		0.89	0.97	0.93	
	CBLS	0.99	1	0.99		–	–	–		0.97	0.96	0.97	
EfficientNetB0[26]	CRM	0.88	0.85	0.86	87.15	0.85	0.88	0.86	85.77	0.85	0.88	0.86	86.45
	CMD	0.53	0.98	0.68		0.95	0.52	0.67		0.95	0.53	0.68	
	CHL	0.83	0.95	0.89		0.91	0.83	0.87		0.93	0.83	0.88	
	CGM	0.99	0.73	0.84		0.72	0.99	0.83		0.72	0.99	0.88	
	CBSD	0.99	0.89	0.93		0.89	0.97	0.93		0.89	0.97	0.83	
	CBLS	1	0.97	0.99		–	–	–		0.97	0.96	0.96	
MobileNetV2[27]	CRM	0.88	0.92	0.9	93.98	0.9	0.88	0.89	92.49	0.9	0.88	0.89	92.86
	CMD	0.83	0.97	0.9		0.96	0.81	0.88		0.97	0.83	0.9	
	CHL	0.93	0.96	0.95		0.95	0.93	0.94		0.94	0.93	0.93	
	CGM	1	0.88	0.93		0.86	0.99	0.92		0.86	1	0.92	
	CBSD	1	0.95	0.97		0.92	1	0.96		0.95	0.97	0.96	
	CBLS	1	0.99	0.99		–	–	–		0.99	0.95	0.97	
RestNet50[28]	CRM	0.96	0.95	0.95	95.59	0.95	0.95	0.95	94.13	0.95	0.96	0.95	95.06
	CMD	0.88	0.99	0.93		0.93	0.84	0.88		0.99	0.84	0.91	
	CHL	0.96	0.96	0.96		0.95	0.93	0.94		0.96	0.96	0.96	
	CGM	0.99	0.94	0.96		0.92	0.97	0.94		0.92	0.97	0.94	
	CBSD	1	0.97	0.99		0.92	1	0.96		0.92	1	0.96	
	CBLS	1	1	1		–	–	–		0.97	0.97	0.97	
NasNetMobile[29]	CRM	0.33	0.97	0.49	66.47	0.97	0.33	0.49	66.81	0.97	0.33	0.49	66.07
	CMD	0.79	0.39	0.53		0.39	0.78	0.52		0.39	0.78	0.52	
	CHL	0.57	0.68	0.62		0.68	0.56	0.62		0.68	0.57	0.62	
	CGM	0.79	0.84	0.82		0.82	0.79	0.81		0.83	0.79	0.81	
	CBSD	0.58	1	0.74		0.93	0.58	0.72		0.59	0.58	0.72	
	CBLS	0.96	0.69	0.81		–	–	–		0.69	0.94	0.81	
Proposed Method	CRM	0.99	0.99	0.99	98.95	0.98	0.99	0.99	97.03	1	0.99	0.99	98.25
	CMD	0.99	0.97	0.97		0.97	0.96	0.96		0.98	0.98	0.98	
	CHL	0.99	0.98	0.98		0.96	0.96	0.96		0.97	0.98	0.97	
	CGM	0.97	0.99	0.99		0.96	0.98	0.98		0.97	0.98	0.98	
	CBSD	0.99	0.99	0.99		0.98	0.95	0.5		0.98	0.98	0.99	
	CBLS	1	1	1		–	–	–		0.99	0.99	0.98	

Abbreviations: CBLS, cassava brown leaf spot; CBSD, cassava brown streak disease; CGM, cassava green mite; CHL, cassava healthy leaf; CMD, cassava mosaic disease; CRM, cassava red mite.

cassava plant leaves. The models are assessed using precision, recall, F1 score, accuracy, and the total number of parameters. The VGG19 model achieved accuracies of 96.59%, 94.08%, and 94.83%, while VGG16 recorded 94.58%, 92.72%, and 93.45%. EfficientNetB0 showed lower accuracies of 87.15%, 85.77%, and 86.45%. MobileNetV2 performed with accuracies of 93.98%, 92.49%, and 92.86%, and ResNet50 achieved 95.59%, 94.13%, and 95.06%. Notably, NasNetMobile yielded considerably lower accuracies at 66.47%, 66.81%, and 66.07%. In contrast, our proposed model outperformed these benchmarks, achieving remarkable accuracies of 98.95%, 97.03%, and 98.25% on all three datasets. Furthermore, our proposed approach has fewer parameters compared to other SOTA models.

A comparison between the suggested model and MobileNetV2 shows that the proposed method is more efficient in terms of accuracy. Similarly, our proposed method is lightweight, with 1.67 million parameters as compared to MobileNetV2's 3.54 million parameters. An analysis of input size and training parameters of the proposed method and SOTA methods is presented in Table 3 and Table 4 presents the performance of the pretrained models. It appears that the pretrained models exhibit good performance while maintaining a minimal false alarm rate. Despite this, the rate of inaccurate predictions remains high and requires enhancement, therefore this research examined the accuracy and inaccuracy of a refined and pre-existing CNN structure combined with an MSA CDDNet. Following these optimizations, CDDNet achieved superior performance compared to other models, exhibiting a reduced number of incorrect predictions and the lowest number of trainable parameters.

The confusion matrices corresponding to the suggested model, which has been trained on three datasets are presented in Fig. 3.



(a) The Cassava Image Dataset (Ramcharan et al., 2017)

(b) Cassava Disease merged (2019-20) (Mwebaze et al., 2019)

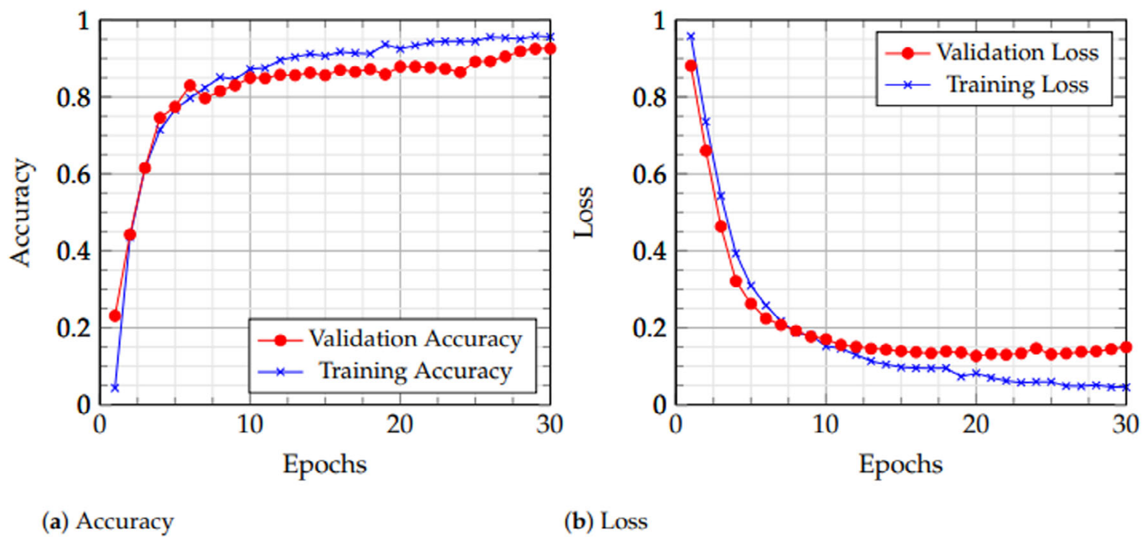
(c) Cassava Leaf Disease Combined

Figure 3. Comparison of confusion matrices for our suggested approach on different datasets. CBLS, cassava brown leaf spot; CBSD, cassava brown streak disease; CGM, cassava green mite; CHL, cassava healthy leaf; CMD, cassava mosaic disease; CRM, cassava red mite.

The deep-blue horizontal diagonals in the Fig. 3 represents TP, whereas the saturation signifies accurate detection. The suggested network demonstrates superior performance in detection as compared to the SOTA models, despite some misclassification of images in all the categories. Figure 4 shows the visualization of the training accuracy and training loss graphs. The horizontal plane depicts the epochs, while the vertical plane represents accuracy and loss. As the number of iterations in the training and validation procedures rises, the line graph representing the accuracy of the model throughout training and validation also changes, as illustrated in Fig. 4(a).

3.3 Qualitative results

To further clarify our comprehension of the proposed model's contributions, we employed explainable AI representation through the utilization of Grad-CAM. Grad-CAM is an effective method that provides insights into the decision-making processes of neural networks by emphasizing important areas inside input images that have a strong impact on the model's predictions. We developed an engaging story using visual representations to illustrate how the attention mechanism boosts the model's comprehension and overall performance. The imagery from Grad-CAM consistently demonstrates the neural network's concentration on detailed and unique features. Such concentration improves the model's skill in identifying complex configurations, textures, and forms, thereby leading to sharper predictions. Utilizing this visual technique not only validates the improvements in precision but also illuminates how the attention mechanism influences the model's decision-making framework. The insights provided by Grad-CAM present a clear and instinctive method for appreciating the efficacy of this approach, effectively showcasing its significance, as illustrated in Fig. 5.



(a) Accuracy (b) Loss

Figure 4. Training, validation accuracy, and loss of our proposed method.

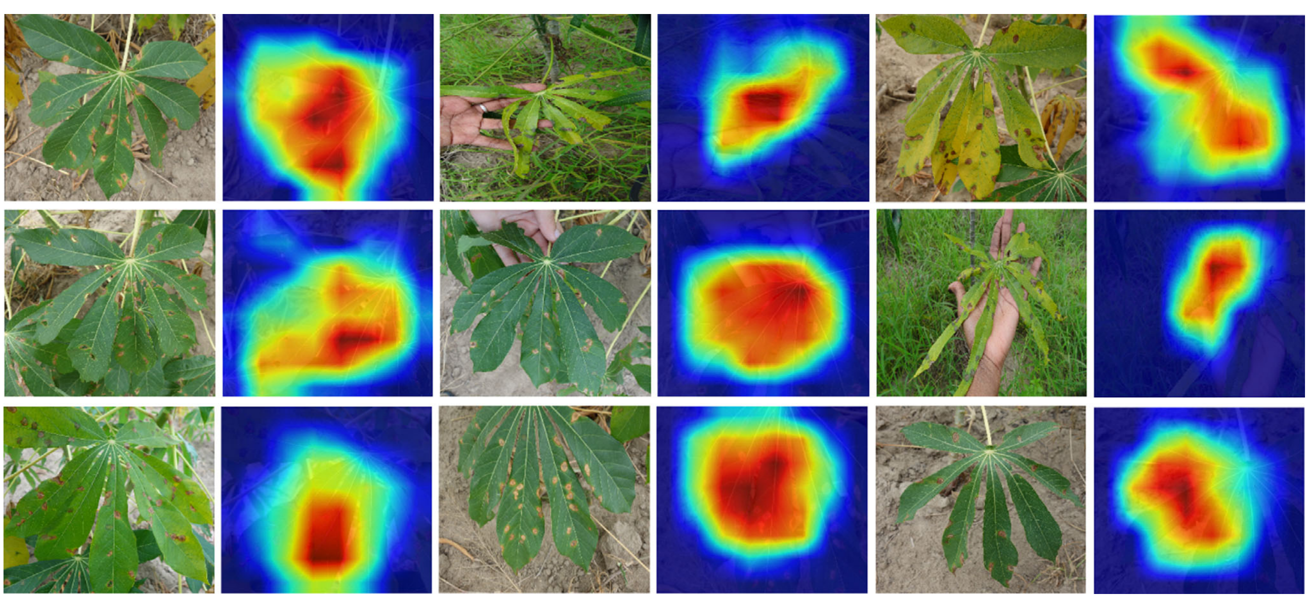


Figure 5. Grad-CAM visualization for real-time detection of cassava disease.

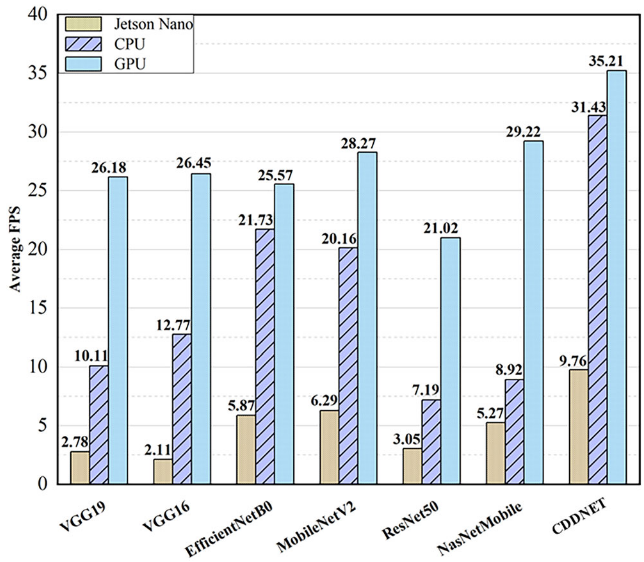


Figure 6. FPS comparison of the proposed method with other SOTA methods.

3.4 Time complexity analysis

Assessing the real-time capabilities across different platforms, including compact edge devices like the Nvidia Jetson Nano, as well as CPU and GPU configurations, is essential for determining the sophistication, efficiency, and deployment feasibility of advanced models. The Nvidia Jetson Nano incorporates a quad-core CPU utilizing ARM Cortex-A57 architecture alongside a discrete GPU leveraging Maxwell architecture. The GPU is equipped with 128 Nvidia CUDA cores and the device has 4 GB of memory. The CPU and GPU details used to analyze the Frames Per Second (FPS) of the proposed CDDNet model are presented. An optimal performance criterion for real-time applications is the achievement of a model that can achieve 30 or more FPS in real-world circumstances. The FPSs for the proposed model using Jetson Nano, CPU, GPU, and Jetson Nano are 31.43, 35.21, and 9.76, respectively. The proposed model is compared to numerous baseline models in terms of the FPS, as presented in Fig. 6. The performance evaluation results demonstrate that while using the Jetson Nano, CPU, and GPU, the VGG19 model achieves frame rates of 2.78, 10.11, and 26.18, respectively. Similarly, the VGG16 model achieves frame rates of 2.11, 12.77, and 26.45, while the EfficientNetB0 model achieves frame rates of 5.87, 21.73, and 28.27. Furthermore, MobileNetV2 achieves 6.29, 20.16, and 28.27 fps, while ResNet50 achieves 3.05, 7.19, and 21.09. Lastly, the NasNetMobile model achieves frame rates of 5.27, 8.92, and 29.22. An analysis of the time complexity of the proposed model in comparison to other baseline models reveals that the results of the proposed model are highly persuasive, therefore the suggested model can process and operate in real time.

4 CONCLUSION

This study introduced an optimal features-assisted lightweight MSA network for the detection of Cassava leaf diseases. The proposed model incorporates an attention module into the MobileNetV3small model. The suggested CDDNet demonstrates high accuracy in the detection of diagnosed plant leaves. It achieves this by effectively analyzing data from Cassava Image Dataset, Cassava Plant Disease Merged (2019–2020) Dataset, and a newly introduced dataset known as Cassava Plant Disease Dataset. The CDDNet method utilizes an end-to-end training architecture to determine and classify deep key points based on their respective classes. In comparison to existing systems, the proposed model outperforms non-attention-based models when it comes to time-effectiveness and classifying Cassava leaf diseases. This demonstrates that the suggested approach for detecting cassava leaf diseases is more robust and generalizable. However, the recommended technique is not suitable in situations where fog or haze leads to reduced visibility, adversely affecting the model's performance. In future, we will focus on improving the resilience and effectiveness of our model by enlarging the dataset by including a broader variety of images. This will involve the incorporation of several forms of interference and a wide array of meteorological factors to replicate real-time fluctuations in the environment. The established augmentation of the dataset is anticipated to greatly enhance the accuracy and performance of the model in various scenarios.

AUTHOR CONTRIBUTIONS

Arailym Dosset: Methodology, conceptualization, investigation, writing – original draft. L. Minh Dang: Project administration, validation. Faisal Alharbi: Investigation. Shabana Habib: Data curation. Nur Alam: Visualization. Han Yong Park: Formal analysis. Hyeonjoon Moon: Resources, supervision, project administration, funding acquisition. All authors have read and agreed to the published version of the manuscript.

ACKNOWLEDGEMENTS

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03038540) and by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries (IPET) through the Digital Breeding Transformation Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (322063-03-1-SB010).

DATA AVAILABILITY STATEMENT

Research data are not shared.

CONFLICT OF INTEREST

The authors have no conflicts of interest.

REFERENCES

1 Mathulaprangsan S and Lanthong K, Cassava leaf disease recognition using convolutional neural networks, in *2021 9th International Conference on Orange Technology (ICOT)*. IEEE, Tainan, Taiwan (2021).
2 Rey C and Vanderschuren H, Cassava mosaic and brown streak diseases: current perspectives and beyond. *Annu Rev Virol* **4**:429–452 (2017).
3 Onyediako P and Adiele J, Enhanced cassava production for food security and economic development in Nigeria: a review. *Niger Agric J* **53**: 204–211 (2022).
4 McCormick D and Schmitz H, *Manual for Value Chain Research on Homeworkers in the Garment Industry*. Mimeo, Institute of Development Studies, Nairobi and Sussex (2001).
5 Food and Agriculture Organization of the United Nations (FAO). *Cassava Diseases in Central, Eastern and Southern Africa (CaCESA)*. (2010).
6 Alabi, O. J., Kumar, P., & Naidu, R. A. *Cassava Mosaic Disease: A Curse to Food Security in subSaharan Africa*. (2011). https://cgspace.cgiar.org/home.

7 Graziosi I, Minato N, Alvarez E, Ngo DT, Hoat TX, Aye TM *et al.*, Emerging pests and diseases of South-East Asian cassava: a comprehensive evaluation of geographic priorities, management options and research needs. *Pest Manag Sci* **72**:1071–1089 (2016).
8 Strange RN and Scott PR, Plant disease: a threat to global food security. *Annu Rev Phytopathol* **43**:83–116 (2005).
9 Devi B, Kumar MN, Chutia M and Bhattacharyya N, Abiotic and biotic stress challenges of cassava in changing climate and strategies to overcome: a review. *Sci Hortic* **305**:111432 (2022).
10 Patil BL and Fauquet CM, Cassava mosaic geminiviruses: actual knowledge and perspectives. *Mol Plant Pathol* **10**:685–701 (2009).
11 Legg J, Somado EA, Barker I, Beach L, Ceballos H, Cuellar W *et al.*, A global alliance declaring war on cassava viruses in Africa. *Food Secur* **6**:231–248 (2014).
12 Ramcharan A, Baranowski K and Ahmed B, Deep learning for image-based cassava disease detection. *Front Plant Sci* **8**:293051 (2017).
13 Sangbamrung I, Praneetpholkrang P and Kanjanawattana S, A novel automatic method for cassava disease classification using deep learning. *J Adv Inf Technol* **11**:241–248 (2020).
14 Sambasivam G and Opiyo GD, A predictive machine learning application in agriculture: cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egypt Inform J* **22**:27–34 (2021).
15 Abayomi-Alli OO, Damaševičius R, Misra S and Maskeliūnas R, Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. *Expert Syst* **38**:e12746 (2021).
16 Riaz SM, Ahsan M and Akram MU, Diagnosis of cassava leaf diseases and classification using deep learning techniques, in *2022 16th International Conference on Open Source Systems and Technologies (ICOSST)*. IEEE, Lahore, Pakistan (2022).
17 Oyewola DO, Dada EG, Misra S and Damaševičius R, Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. *PeerJ Comput Sci* **7**:e352 (2021).
18 Hassan SM and Maji AK, Plant disease identification using a novel convolutional neural network. *IEEE Access* **10**:5390–5401 (2022).
19 Liu M and Liang H, Research on cassava disease classification using the multi-scale fusion model based on EfficientNet and attention mechanism. *Front Plant Sci* **13**:1088531 (2022).
20 Ravi V, Acharya V and Pham TD, Attention deep learning-based large-scale learning classifier for cassava leaf disease classification. *Expert Syst* **39**:e12862 (2022).
21 Maryum A, Akram MU and Salam AA, 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)*. IEEE, Karachi, Pakistan (2021).
22 Mehta S, Kukreja V and Gupta A, Revolutionizing cassava leaf disease detection with federated learning CNN and multi-level severity assessment, in *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*. IEEE, Kollam, India (2023).
23 Lilhore UK, Imoize AL, Lee C-C, Simaiya S, Pani SK, Goyal N *et al.*, Enhanced convolutional neural network model for cassava leaf disease identification and classification. *Mathematics* **10**:580 (2022).
24 Maimon O and Rokach L, *Decomposition Methodology for Knowledge Discovery and Data Mining*. Springer, Berlin (2005). https://link.springer.com/book/10.1007/b107408.
25 Simonyan K and Zisserman A, Very deep convolutional networks for large-scale image recognition. *arXiv* (2023).
26 Tan M and Le Q, Efficientnet: rethinking model scaling for convolutional neural networks, in *International Conference on Machine Learning*. PMLR, California, USA (2019).
27 Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L-C, Mobilenetv2: inverted residuals and linear bottlenecks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Salt Lake City, USA (2018).
28 He K, Zhang X, Ren S and Sun J, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, USA (2016).
29 Zoph B, Vasudevan V, Shlens J and Le QV, Learning transferable architectures for scalable image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Salt Lake City, USA (2018).
30 Mwebaze E, Gebru T, Frome A, Nsumba S and Tusubira J, iCassava 2019 fine-grained visual categorization challenge. arXiv preprint arXiv: 1908.02900 (2019).