



A deep feature fusion network using residual channel shuffled attention for cassava leaf disease detection

R. Karthik¹ · R. Menaka¹ · M. V. Siddharth² · Sameeha Hussain³ · Bala Murugan³ · Daehan Won⁴

Received: 28 November 2022 / Accepted: 1 August 2023

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

Cassava is a significant source of carbohydrates for tropical populations. However, diseases caused by agents such as bacteria, viruses, fungi, and phytoplasmas cause considerable economic damage to these crops. Existing methods for cassava disease detection require farmers to seek the assistance of agricultural experts for visual inspection and diagnosis, which is challenging and laborious. Most studies have employed pre-trained convolutional neural networks to detect diseases in cassava leaves. Also, it is essential to design customized deep neural networks specific to the target domain for precise classification. This research proposes a novel deep fusion of two networks, residual channel shuffled attention network and Efficientnet. The first network, RCSANet, was presented to capture contextual information using depthwise separable convolution effectively. It also integrates significant inter-spatial and inter-channel information using the triplet attention module and employs shuffled group convolution to capture features from distinct filter groups. As a result of incorporating the above architectural enhancements, the proposed feature fusion network exhibited better performance than the existing studies. The proposed network was trained on the Kaggle cassava leaf disease dataset with 21,367 samples and yielded a classification accuracy of 93.25%.

Keywords Plant disease · Hybrid features · Multi-path network · Attention · Deep learning

1 Introduction

Cassava was originally introduced into Africa by Portuguese trade merchants in the sixteenth century. It currently serves as the most important tropical root crop and the highest producer of carbohydrates among staple crops. The United Nations Food and Agriculture Organization ranks cassava fourth among developing-country food crops, after rice, maize, and wheat. Cassava's starchy roots provide a significant amount of energy, while the leaves are palatable and high in protein content. Additionally, cassava is increasingly used as animal feed and in the production of industrial products such as alcohol, sweeteners, and starches for textiles and paper. Cassava's ability to be preserved in the ground throughout seasons makes it a valuable backup food source in case of crop failure [2]. However, it is highly susceptible to viral, fungal, and bacterial pathogens that can significantly impact yield losses [17]. Expanding our knowledge of cassava diseases and finding effective methods to curb them is crucial to maximizing cassava yields. Cassava bacterial blight is a

✉ R. Karthik
r.karthik@vit.ac.in

R. Menaka
menaka.r@vit.ac.in

M. V. Siddharth
siddharth.mv2019@vitstudent.ac.in

Sameeha Hussain
sameeha.hussain2019@vitstudent.ac.in

Bala Murugan
balamurugan.p2019@vitstudent.ac.in

Daehan Won
dhwon@binghamton.edu

¹ Centre for Cyber Physical Systems, Vellore Institute of Technology, Chennai, India

² School of Mechanical Engineering, Vellore Institute of Technology, Chennai, India

³ School of Electronics Engineering, Vellore Institute of Technology, Chennai, India

⁴ System Sciences and Industrial Engineering, Binghamton University, Binghamton, USA

particularly devastating disease on a global scale, which often results in the complete loss of crops under certain conditions [7]. Other diseases, such as Cassava mosaic and *Cercospora* leaf spots, are also significant limiting factors in cassava production, while several root rots and *Phylosticta* leaf spots can cause substantial yield losses in certain environmental conditions [14]. Some of the widely prevalent diseases that affect cassava crops include cassava mosaic disease (CMD), brown leaf spot (BLS), cassava brown disease streak (CBSD), cassava red mite damage (RMD), and cassava green mite damage (GMD). Out of these diseases, CMD and CBSD are the most significant impediments to cassava production and food security, which result in nearly \$1 billion in annual losses [34].

Detecting the underlying disease affecting the crop is critical to fully harnessing cassava's agricultural potential. Currently, existing disease detection and identification methods require farmers to seek the assistance of agricultural experts for visual inspection and diagnosis, which is inefficient, time-consuming, and labor-intensive [22]. To overcome these challenges, an intelligent system is required, which paves the way for the implementation of computer-aided diagnosis (CAD). Advances in artificial intelligence have led to more efficient and powerful CAD systems. This study aims to automate the process of disease detection using deep learning techniques, which aids farmers to potentially save their crops and the country's economy.

2 Related works

This section analyzes numerous deep learning-based approaches applied to cassava disease detection.

Deep neural networks can incrementally learn the essential features through their hidden layers. Moreover, their flexibility lets them adapt to new problems in the future. One of the most significant advantages of deep learning is its ability to automate feature extraction and optimally tune the hyperparameters for the desired outcome.

In recent years, several studies have focused on plant disease classification to foster efficient plant disease detection solutions for farmers. Plant-based disease detection that uses deep learning techniques has become a prominent research domain. Convolution neural networks (CNN) have been used extensively in the majority of studies. Several architectures such as MobilenetV2 [1, 3, 31], Resnet [18, 23], InceptionV3 [3, 18, 26], Densenet [16], and Efficientnet [5, 9, 15, 19, 28, 36] have been employed for classification.

A significant number of studies have employed transfer learning-based approaches for plant disease detection.

Abayomi et al. [1] proposed a method that employed a pre-trained MobileNetV2 architecture for cassava disease detection. Moreover, data augmentation was performed on low-quality images to improve the performance of the network. Ayu et al. [3] presented another approach where they utilized a MobileNetV2 model for cassava disease classification. Ramcharan et al. [26] proposed a methodology where they employed Inception V3 for feature extraction. Furthermore, support vector machine (SVM) and k-nearest neighbors (KNN) were employed for the classification of cassava diseases. Few transfer learning-based approaches using the Efficientnet architecture were reported in the literature for the cassava disease classification. Ravi et al. [28] presented an approach where a pre-trained Efficientnet model was utilized for effective feature extraction. Furthermore, random forest and SVM were employed for classification. Similar approaches using a pre-trained Efficientnet architecture were proposed for cassava disease detection [5, 9, 15, 19, 36]. Apart from using only one architecture, many studies compared the performance of two or more models for the classification of cassava diseases. Mathulapransan et al. [16] performed a comparison of current state-of-the-art CNN models. It could be inferred that the Densenet-121 model with brightness augmentation outperformed the other architectures. In another similar approach, Megha et al. [18] presented a comparison between three pre-trained architectures, namely InceptionResnetV2, InceptionV3, and Resnet-50. Here, InceptionResnetV2 resulted in greater overall classification performance. Metlek [20] presented an approach where Resnet-50 and MobileNetV2 architectures were compared for feature extraction. Out of the two models, Resnet-50 with SVM performed better in terms of overall classification accuracy.

Chen et al. [5] formulated a new loss function to treat noisy labels present in the cassava image dataset. Here, various CNN models were compared against each other, and Resnet-50 resulted in greater classification accuracy and performance. Gao et al. [9] proposed a methodology that aims to monitor cassava diseases based on the use of HSV color space to preprocess the diseased leaves images. Moreover, the Efficientnet architecture was used to perform cassava disease classification. Thai et al. [33] applied a pre-trained vision transformer model to identify infected leaves. The model was further quantized to reduce the computational cost for compatibility on mobile devices. Few studies have employed an object detection model for the effective detection of diseases from images. Ramcharan et al. [27] proposed one such approach in which a single-shot detector (SSD) model was applied using a pre-trained MobileNet as the backbone. In another similar approach, Megha et al. [30] applied a Faster R-CNN model for disease detection that uses bounding boxes in cassava leaf

images. Furthermore, a CNN model was used for binary classification.

Apart from pre-trained architectures, a few custom CNN architectures were also presented for cassava disease classification. Sambasivam et al. [29] proposed a method to counter class imbalance using techniques such as class weight and synthetic minority oversampling technique (SMOTE). Furthermore, they employed focal loss with a custom deep CNN for the detection of cassava disease. In another approach, Oyewola et al. [23] employed a custom CNN architecture with residual connections for cassava disease classification. The results demonstrate that the network with the residual connection outperforms a simple 3-layer CNN model. Hassan et al. [10] presented an approach where a modified CNN based on Inception-V3 architecture was applied for cassava disease identification.

In addition to classification, many of the studies performed segmentation for the effective detection of cassava disease. A two-stage approach for cassava disease detection was presented by Maryum et al. [15]. Image segmentation was performed by an Unet architecture to remove background noise, followed by feature extraction and classification using a pre-trained Efficientnet-B4 model. Another approach using segmentation was proposed by Patike et al. [24] where the architecture was modified to replace classical convolution layers with depthwise separable convolution layers Table 1.

3 Research gaps and motivation

The proposed study effectively addresses the following research gaps in cassava leaf disease detection:

1. The cassava dataset employed in the existing studies suffers from severe class imbalance. This imbalance will have a detrimental effect on the performance of the model because it would not be able to learn significant feature patterns.
2. While most of the existing studies in cassava disease detection employ standard CNN architectures and transfer learning methods, customized architectural design specific to the input data will result in better generalization of the trained model.
3. Existing studies give equal weightage to all channels regardless of their importance. They do not provide precise weightage to a single feature map both within and across channels. Moreover, contextual information is necessary to extract correlations between neighboring pixels.

4 Research contributions

The following are the contributions made toward addressing the gaps stated above:

1. To tackle the problem of class imbalance, focal loss was employed in the proposed method. Moreover, extensive data augmentation is performed to introduce variation to improve feature learning capability and also to prevent overfitting.
2. The proposed CNN architecture integrates various convolutional layers, such as depthwise separable convolution and shuffled group convolution, to learn robust features across the network. Moreover, this design improves information flow because variably sized receptive fields are utilized at the inception of the network.
3. The proposed study utilizes a residual channel shuffled attention (RCSA) block to extract features from the images effectively. The triplet attention block in the residual connection was used to extract inter-channel and inter-spatial features adaptively. Furthermore, separate branches that comprise depthwise separable convolution and average pooling are added to extract the relationship between nearby pixels effectively. Therefore, through the addition of these specific blocks, the overall feature representation power and gradient flow of the network is improved.
4. To increase the overall performance and generalizability of the network, multi-level feature fusion is performed in this work. Here, the features from the Efficientnet model are concatenated with the features of the proposed architecture for precise classification.

5 Proposed system

The proposed methodology consists of a fusion of two CNN architectures: (1) residual channel shuffled attention (RCSA) network and (2) Efficientnet-B0. The motivation behind the RCSA network is to extract predominant features with the use of residual links embedded with an attention module. Moreover, multiple paths are employed for the aggregation of features across the network effectively. Finally, the features from the RCSA network and Efficientnet are fused to improve the overall performance and generalization capability [8]. First, the data are loaded and split into a training set and a validation set. The images are reduced to dimensions of 448×448 and are passed into the train and validation loaders, where online augmentation is performed. A high-level workflow of the proposed methodology is presented in Fig. 1.

Table 1 A summary highlighting the strengths and weaknesses of existing works on cassava disease detection

S. No.	Source	Method	Strengths	Limitations
1	Ramcharan et al. [26]	Employed Inception V3 for feature extraction. Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) were used for the classification of cassava diseases	Employed a pre-trained hybrid CNN for efficient classification	Although transfer learning is effective, different initial and target domains may lead to negative transfer
2	Ramcharan et al. [27]	A single Shot Detector (SSD) model was applied using a pre-trained MobileNet as the backbone	Employed an object detection model with a lightweight backbone architecture for deployment on edge devices	Although transfer learning is effective, different initial and target domains may lead to negative transfer
3	Sangbamrung et al. [30]	Applied a Faster R-CNN model for disease detection using bounding boxes in cassava leaf images A CNN model was used for binary classification	Utilized two separate models for object detection and classification	Multi-class classification may help to provide more elaborate results
4	Surya et al. [31]	MobileNetV2 model for cassava disease classification	Lightweight CNN model that reduces complexity cost with comparable performance	Although transfer learning is effective, different initial and target domains may lead to negative transfer
5	Abayomi et al. [1]	Pre-trained MobileNetV2 architecture for cassava disease detection Data augmentation was performed on low-quality images to improve the performance of the network	Proposed an improved classification model based on data augmentation to handle low-quality images	Different initial and target domains may cause hindrances to the model performance when transfer learning is employed
6	Gao et al. [9]	Proposed a methodology that aims to monitor cassava diseases based on the use of HSV color space to preprocess the diseased leaves images Efficientnet architecture was employed to perform cassava disease classification	HSV color space was used for image preprocessing to reduce the loss of information during the final classification	Hardware limitations have been reported for the proposed system
7	Maryum et al. [15]	Image segmentation was performed by an Unet architecture to remove background noise followed by feature extraction and classification using a pre-trained Efficientnet-B4 model	Performed segmentation to remove irrelevant features such as background noise	Class imbalance in the dataset was not addressed
8	Methil et al. [19]	Proposed a One-vs-All methodology to perform multi-class classification. Efficientnet-B4 was used as the backbone architecture	Performed multi-class classification using a novel methodology for effective and precise classification	Different initial and target domains may cause hindrances to the model performance when transfer learning is employed
9	Metlek [20]	Resnet-50 and MobileNetV2 architectures were compared for feature extraction Resnet-50 with SVM performed better in terms of overall classification accuracy	Background segmentation has been performed for all the images in the dataset. Furthermore, classification was done by pre-trained models	Although transfer learning is effective, different initial and target domains may lead to negative transfer
10	Oyewola et al. [23]	Employed a custom CNN architecture with residual connections for cassava disease classification	Used residual connections for improved gradient flow across the network	The model fitting has been reported. Gamma correction may not be the ideal technique for image enhancement
11	Patike et al. [24]	The architecture was modified to replace classical convolution layers with depthwise separable convolution layers	Regular convolutional layers have been replaced with separable convolutions to reduce computational complexities	Class imbalance in the dataset was not addressed
12	Ravi et al. [28]	The pre-trained Efficientnet model was utilized for effective feature extraction Random Forest and SVM were also employed for classification	Used an ensemble of pre-trained Efficientnet architectures	The proposed method is sensitive to imbalanced data

Table 1 (continued)

S. No.	Source	Method	Strengths	Limitations
13	Sambasivam et al. [29]	Solved class imbalance using techniques such as class weight and Synthetic Minority Oversampling Technique (SMOTE) Furthermore, employed a custom deep CNN for the detection of cassava disease	Addressed class imbalance using different techniques and used a simple 3-layer CNN for fast classification	Attention layers are necessary to extract salient features for precise classification
14	Thai et al. [33]	Applied a pre-trained Vision Transformer model to identify infected leaves	The model was further quantized to reduce the computational cost for compatibility on mobile devices	Different initial and target domains may cause hindrances to the model performance when transfer learning is employed
15	Chen et al. [5]	Various CNN models were compared against each other, and Resnet-50 resulted in greater classification accuracy and performance	Formulated a new loss function to treat noisy labels present in the cassava image dataset	Neglects data distribution of individual datasets, which may lead to inaccurate classification
16	Hassan et al. [10]	Modified CNN based on Inception-V3 architecture was applied for cassava disease identification	Employed residual connections and modified inception blocks in a custom CNN for efficient classification	Attention layers are necessary to extract salient features for precise classification
17	Vijayalata et al. [36]	Pre-trained Efficientnet-B0 architecture was utilized for the early detection of cassava disease	Utilized a pre-trained model that provides the best trade-off between computational complexity and performance	Although transfer learning is effective, different initial and target domains may lead to negative transfer

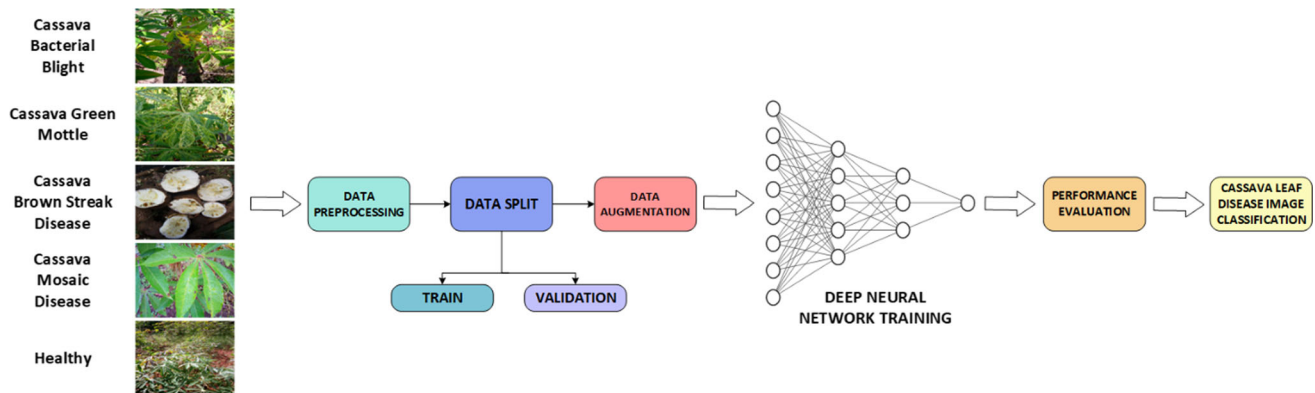


Fig. 1 Overall workflow of the proposed methodology based on a custom convolutional network. The images from the dataset are split into a training set and validation set before being fed into the neural

network for feature learning. Online augmentation is performed using the Albumentations [4] library

5.1 Proposed network

The proposed system is a feature fusion network that comprises two models, namely RCSANet and Efficientnet-B0, as presented in Fig. 2. The extracted features from both the individual networks are fused to produce greater performance and stability. Both networks contain attention mechanisms that have proved to be helpful in identifying salient regions in the image [37].

The Efficientnet architecture achieves state-of-the-art results with a low computational overhead due to the

property of compound model scaling [32]. The squeeze-and-excitation (SE) block in the model focuses on extracting channel features adaptively across the network, whereas the triplet attention module in the RCSANet focuses on learning contextual information as well as inter-spatial and inter-channel features [21]. Therefore, both networks support each other for effective feature learning.

5.1.1 Residual channel shuffled attention network

The architecture of the proposed RCSA network is presented in Fig. 2. The CNN operates with the input

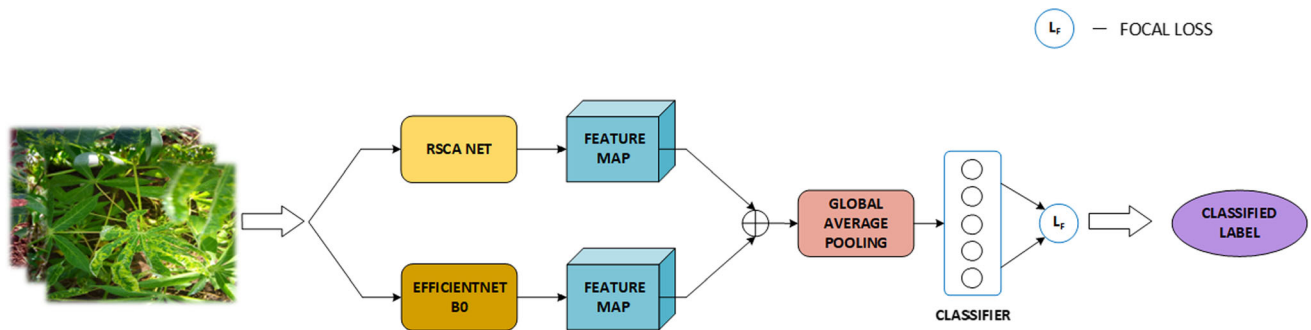


Fig. 2 Schematic diagram of the hybrid dual-path feature fusion network that consists of an ensemble of RCSANet and Efficientnet. The feature maps from both networks are concatenated before passing

through the global average pooling layer for dimensionality reduction and classification

dimensions of 448×448 . The feature depth of the CNN grows by a factor of 2 as it progresses through the layers. The input images are propagated across multiple convolution and attention layers before being fed to the classifier. Depthwise separable convolutions and shuffled grouped convolutions are employed to reduce the computational complexity of the model. The advantages of these convolutions are as follows: (1) The correlations between features and contextual information are enhanced; (2) they were included to improve the flow of gradients between adjacent layers; (3) they reduce the cost of computation (learnable parameters) while still being able to demonstrate comparable performance; (4) residual links enable the

aggregation of multi-scale features, which improves the learning capabilities of the network [11].

The proposed network is composed mainly of two RSCA blocks for effective feature extraction. The RSCA blocks extract multi-scale attention-guided features with the help of different convolution layers and attention layers. In general, the convolution layers in the linear path extract mainline features, whereas the depthwise separable convolution and average pooling paths (*side_path1* and *side_path2*) act as auxiliary layers that provide contextual information for improved feature learning.

Initially, the network forks into mainline and sideline paths, as shown in Fig. 3. The mainline path consists of a

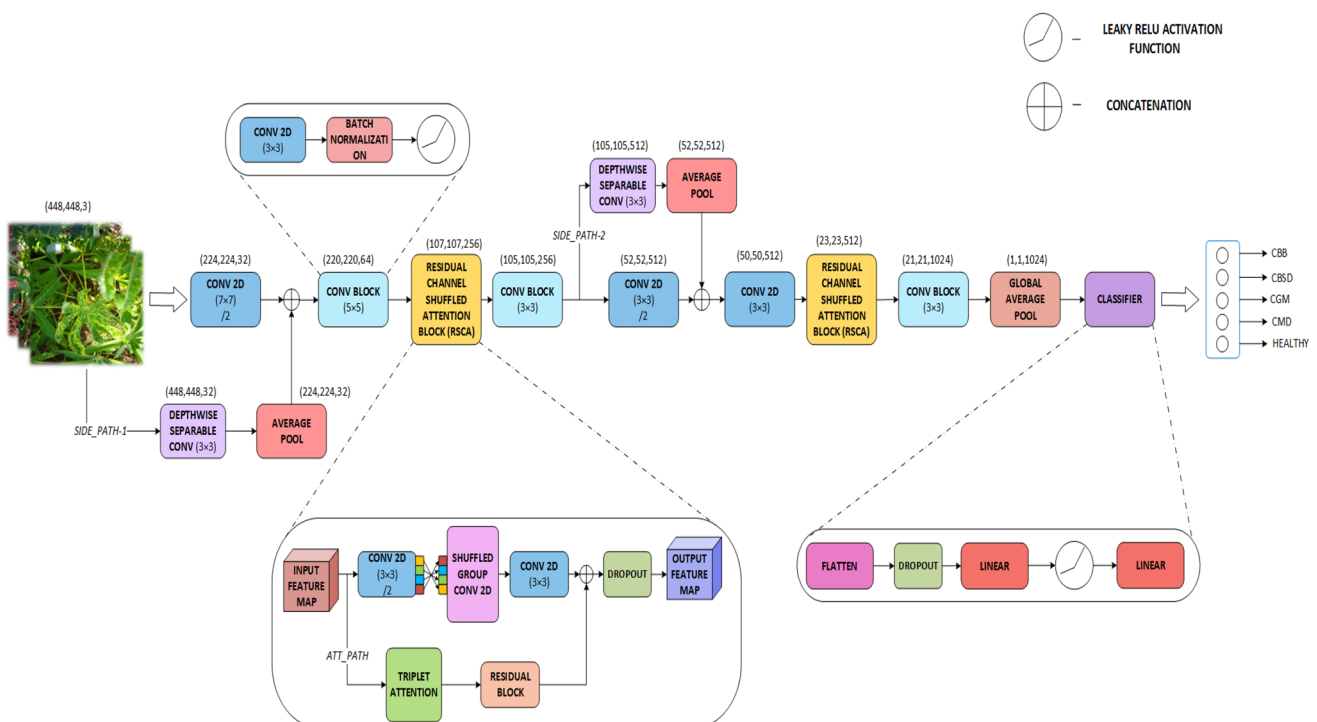


Fig. 3 Illustration of the proposed residual channel shuffled attention network for effective extraction of multi-scale features with the help of auxiliary paths and attention modules. Moreover, the group shuffle layer enhances the feature learning capability of the network

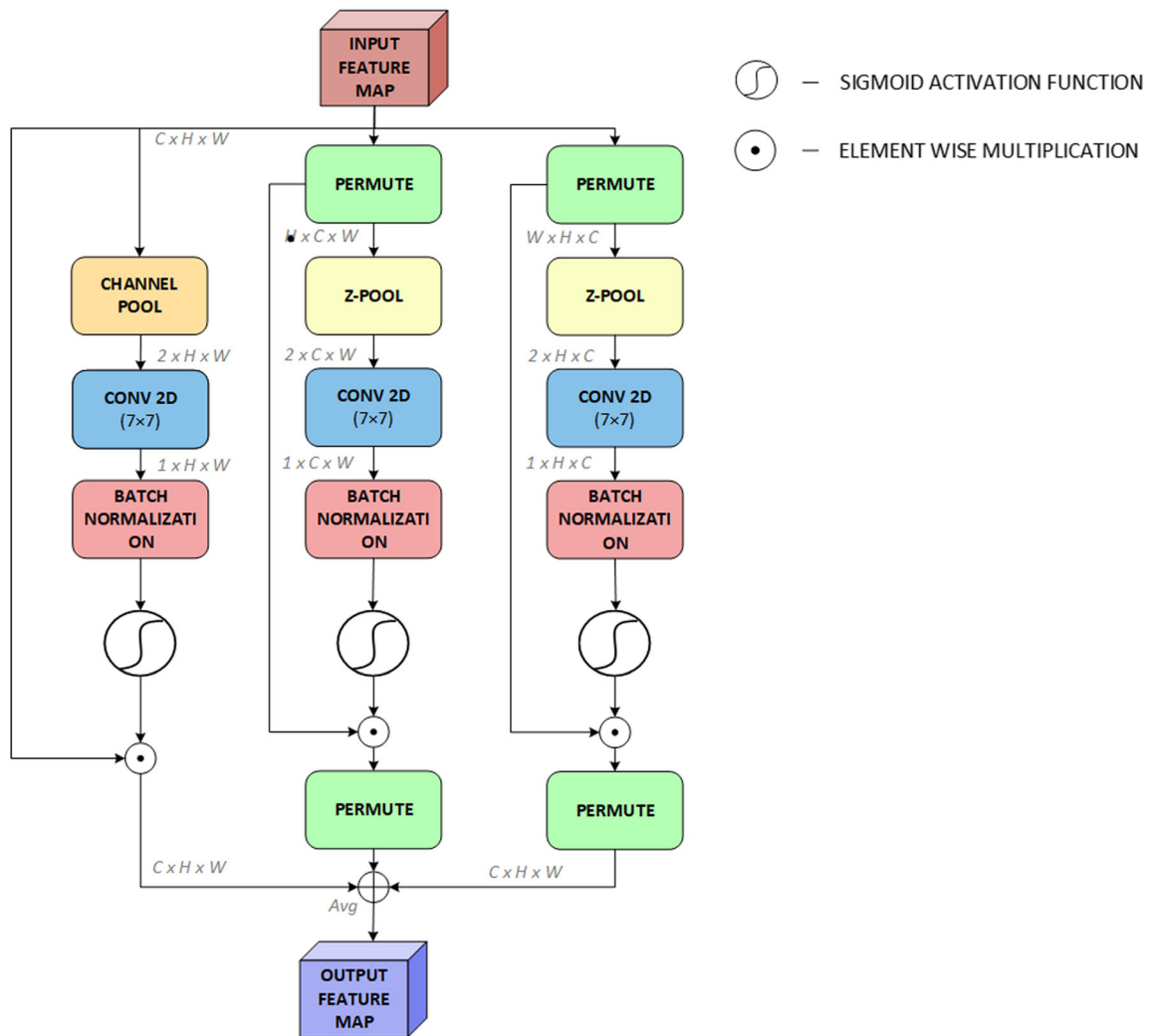


Fig. 4 Schematic diagram of the triplet attention module. The rightmost and middle branches are responsible for computing attention weights across the different channels and spatial dimensions.

convolution layer with kernel size 7×7 for low-level feature extraction. Large kernel filters at the beginning of the network allow for more expressive power and are responsible for extracting maximum spatial information from the cassava images directly. The sideline path consists of a depthwise separable convolution followed by an average pooling layer. Depthwise separable convolutions are less computationally expensive than regular spatial convolutions because it handles spatial and depth dimensions separately. This was included to improve the extraction of contextual information, which is often overlooked. Moreover, dense pixel connectivity is induced by establishing correlations from all the feature channels. The *side_path1* fuses with the mainline path after the 7×7 convolution layer. These features are concatenated and filtered by the following convolution block. It consists of a

The leftmost branch captures spatial dependencies across the network. Finally, the weights from the three branches are aggregated using the concatenation operation

convolution layer with a kernel size of 5×5 , batch normalization, and a Leaky Relu activation function.

The features were then propagated to the RCSA block for effective feature learning. Here, the feature maps from the convolution layer are acted upon by the shuffled group convolution layer. Shuffled group convolution is included to improve information and gradient flow across the channels in the network by learning salient features from distinct filter groups in the input feature map [38]. Furthermore, the residual link *att_path* consists of a triplet attention module followed by a residual block. The triplet attention module illustrated in Fig. 4 extracts both inter-spatial and inter-channel information. It was included to strengthen the power of the spatial and channel attention mechanism by utilizing cross-dimensional interaction. From an input tensor $X \in \mathbb{R}^{C \times H \times W}$, the final refined attention map y obtained can be represented by Eq. (1):

$$y = \frac{1}{3}(\overline{X_1\omega_1} + \overline{X_2\omega_2} + X_3\omega_3) \quad (1)$$

where ω_1 , ω_2 and ω_3 represent the cross-dimensional attention weights computed from the three branches of triplet attention. Here, X_1 , X_2 are tensors reduced by the branches with Z-pool and X_3 is the tensor generated by the third branch. The Z-pool layer can be mathematically represented by Eq. (2):

$$Z - \text{pool}(X) = \text{MaxPool}_d(X) \oplus \text{AvgPool}_d(X) \quad (2)$$

where d is the dimensions across which the max pooling and average pooling operations take place, \parallel is the concatenation operation, and X is the input tensor.

Therefore, the network learns to focus on predominant features, and gradient flow is improved with negligible computational overhead. The output feature map was fed to the residual block, which fuses with the mainline path. Overall, the residual link improves feature learning capability by yielding multiple network paths for information flow and also prevents the problem of vanishing gradients. Dropout was applied to provide a regularization effect to prevent overfitting. The output feature map was propagated to a convolution block. The path splits into two branches again where the feature map passes through *side_path2*, a track identical to *side_path1* that consists of depthwise separable convolution and average pooling layers. The features from the *side_path2* were fused to the mainline path, which is further propagated to a second RCSA block. The output feature map was then fed to a final convolution block for effective feature extraction.

5.1.2 Classification

In the case of RCSANet, a global average pooling layer was applied to reduce the dimensions of the feature maps with a low computational cost. It is reported to prevent overfitting and is also more robust to spatial translations of the input feature map [13]. The reduced feature map was fed to a flattened layer followed by a dropout layer. Furthermore, the features were passed on to fully connected layers for the classification of cassava disease. The RCSANet consists of 20,149,575 parameters.

For the effective classification of the proposed hybrid dual-path feature fusion network, global average pooling was applied to reduce the dimensions of the feature maps. The reduced feature maps were propagated to the final fully connected layer for the classification of cassava leaf diseases into five classes, namely CBB, CBSB, CGM, CMD, and healthy. The hybrid dual-path network consists of 23,163,528 trainable parameters.

The focal loss was adopted to measure the performance of the deep network to classify the five targets. It aims to

solve the problem of class imbalance by downweighing the easy samples and focusing on the hard examples. The network parameters were learned by optimizing the focal loss of the predicted class probabilities with the target class given by Eqs. (3) and (4):

$$\text{FL}(p_t) = -\alpha_t(1 - p_t)^\gamma \log p_t \quad (3)$$

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{otherwise} \end{cases} \quad (4)$$

where $(1 - p_t)^\gamma$ is the modulating factor, γ is the focusing parameter, α_t are the class weights, and p is the class probability. As a result, the focal loss is a dynamically scaled cross-entropy loss where the scaling factor decays to zero as confidence in the correct class increases. It offers greater performance than the regular cross-entropy loss, especially in problems that involve severe class imbalance.

6 Results

This section presents the dataset description, data preprocessing and data augmentation, environmental setup, ablation studies, and performance analysis.

6.1 Dataset description

The proposed model of this research work was developed using the Cassava Leaf Disease Classification dataset sourced from [12]. The five main classes include cassava bacterial blight (CBB), cassava mosaic disease (CMD), cassava brown streak disease (CBSD), cassava green mite (CGM), and cassava healthy leaf, as illustrated in Fig. 5, Table 2.

6.2 Data preprocessing and data augmentation

The different techniques used for data preprocessing and data augmentation are discussed in this subsection. The cassava dataset consists of a total of 21,367 images, with each image having a dimension of 800×600 pixels. Before training, the images were resized to 448×448 pixels for effective feature extraction. The dataset was split into a training set and a validation set in the ratio of 80:20.

Because the dataset contains a majority of samples that belong to CMD, the class imbalance was observed, as reported in Fig. 6. This may result in the inferior performance of the proposed network. To overcome class imbalance, it is necessary to perform data augmentation. This also enables the model to learn more robust features and better generalize. The following augmentations were carried out using the Albumentations [4] library: (1) random portions of an image were cropped and resized to a

Fig. 5 Examples of different classes from the cassava dataset
a CBB, **b** CBSD, **c** CGM,
d CMD

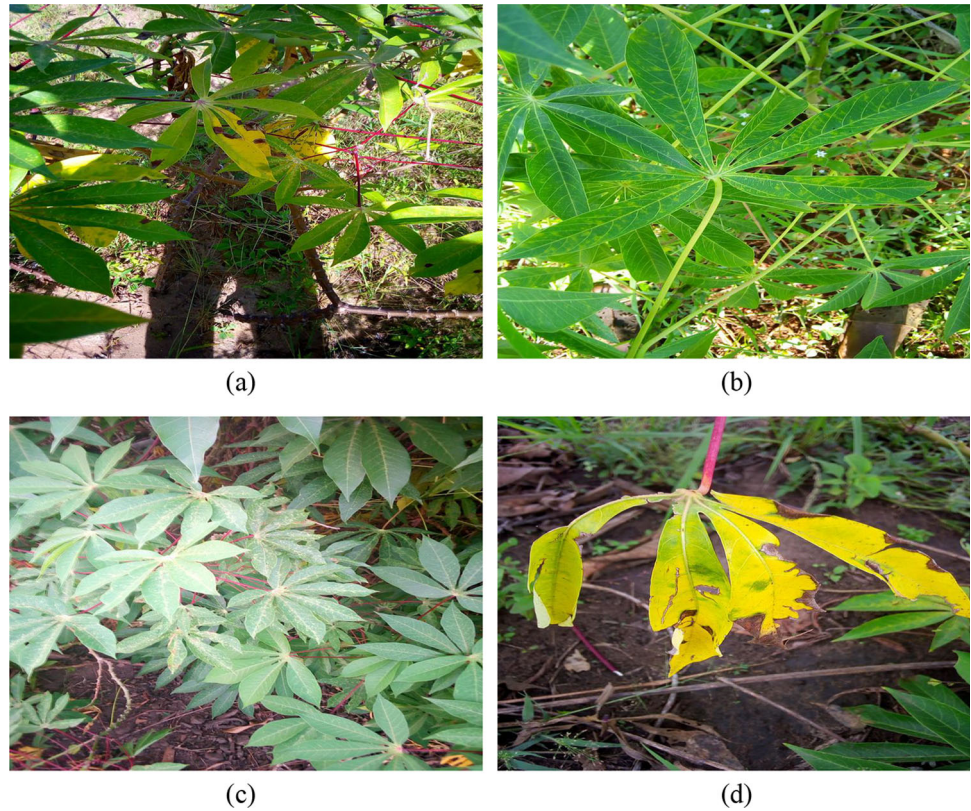


Table 2 Details of Cassava dataset. More than 50% of images belong to class CMD which results in a severe class imbalance that can affect model training and performance

Class	Number of images
Cassava bacterial blight (CBB)	1087
Cassava brown streak disease (CBSD)	2189
Cassava green mottle (CGM)	2386
Cassava mosaic disease (CMD)	13,158
Healthy	2577
Total	21,367

given size; 2) randomly transposing the input images; 3) randomly flipping the images horizontally and vertically; 4) random affine transformations were applied on input images using ShiftScaleRotate; 5) random change of hue, saturation, and value of the input images; 6) randomly varied the brightness and contrast of the images; and 7) coarse dropout and cutout augmentation were applied to remove rectangular regions in the input images randomly.

6.3 Environmental setup

The proposed network was trained on a 24 GB Nvidia A10G GPU. The deep learning model was implemented

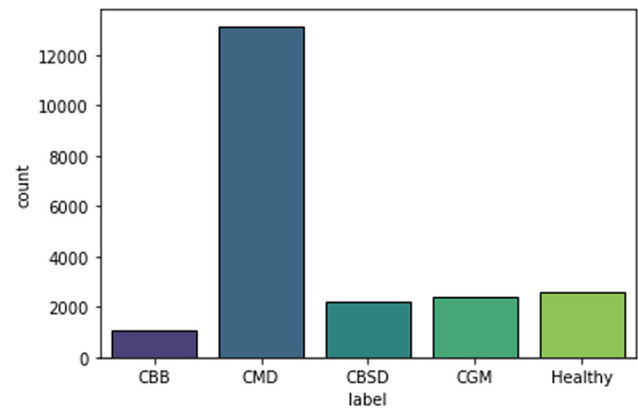


Fig. 6 Distribution of the five classes in the dataset

using Pytorch on an AWS EC2 instance. The system specifications are Ubuntu 20.04, 4 AMD vCPUs, and 16 GB RAM. To optimize the parameters effectively, both stochastic gradient descent (SGD) and adaptive moment estimation (Adam) were tested. It was found that Adam yielded better results over SGD with an initial learning rate of 0.0001 and a weight decay of 0.000001. Moreover, cosine annealing with warm restarts was employed as the learning rate scheduler to prevent the learner from getting stuck in a local minimum. To increase the performance of the network while reducing the training time, mixed-

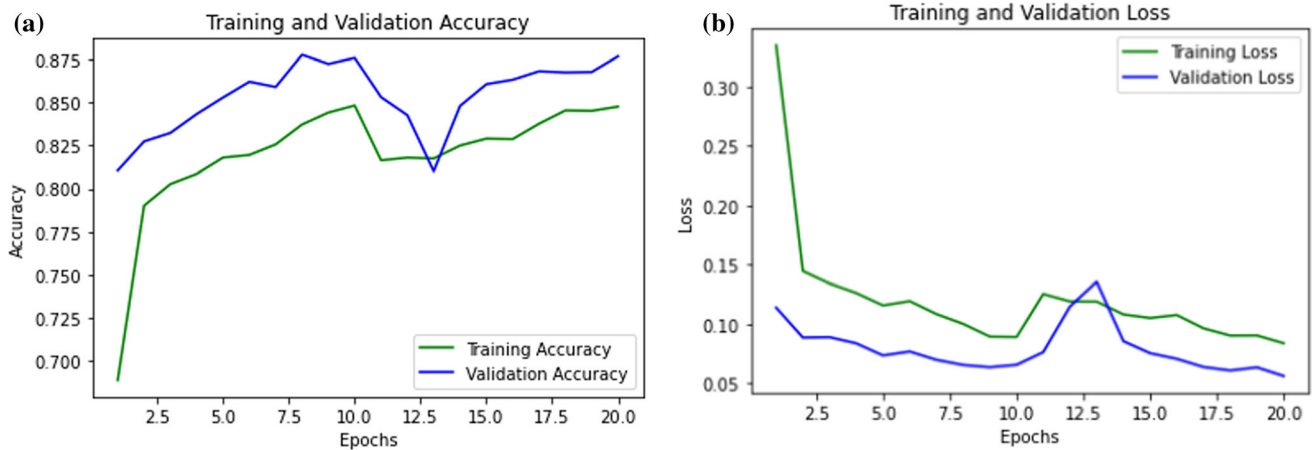


Fig. 7 Analysis of Efficientnet-B0 backbone architecture: **a** accuracy and **b** loss. The pre-trained model reached stable performance in 20 epochs

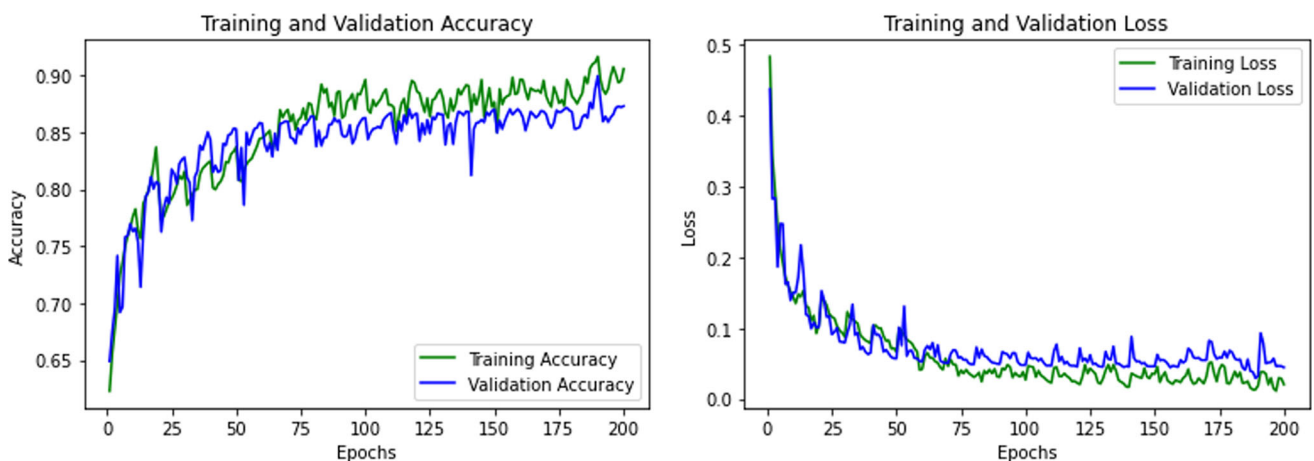


Fig. 8 Accuracy and loss plots for the analysis of RCSANet architecture. The proposed custom CNN was trained and validated for 200 epochs

precision training using the AMP CUDA library was utilized. The torch datatypes are converted to half-precision (FP16) format, the default format being float-point 64 (FP64). Additionally, gradient scaling was used during backpropagation to prevent the gradients from becoming zero. Overall, the time taken to train the neural network was reduced by a factor of 3 without compromising performance.

6.4 Ablation studies

Ablation experiments were performed to validate the effectiveness of the various components of the proposed architecture. The CNN was incrementally trained with the addition of a new component. The degree of performance improvement is measured by the following metrics: (1) accuracy, (2) F1-score, (3) precision, and (4) recall.

6.4.1 Analysis of Efficientnet-B0

This subsection analyzes the performance of the Efficientnet model pre-trained on imagenet weights. The baseline model was trained for 20 epochs, where each epoch took nearly two minutes to complete. The observations are presented in Fig. 7. An accuracy of 87.7% was obtained, and the total number of trainable parameters in the Efficientnet-B0 model was 4,013,953.

6.4.2 Analysis of residual channel shuffled attention network (RCSANet)

The proposed RCSANet was trained on the cassava leaf disease dataset for 200 epochs, where each epoch took nearly five minutes to complete. The RCSA block in the network plays a major role in the extraction of features. It adaptively learns robust spatial and channel features to weigh significant group features across the blocks. The presence of a shuffled group convolution layer enhances feature

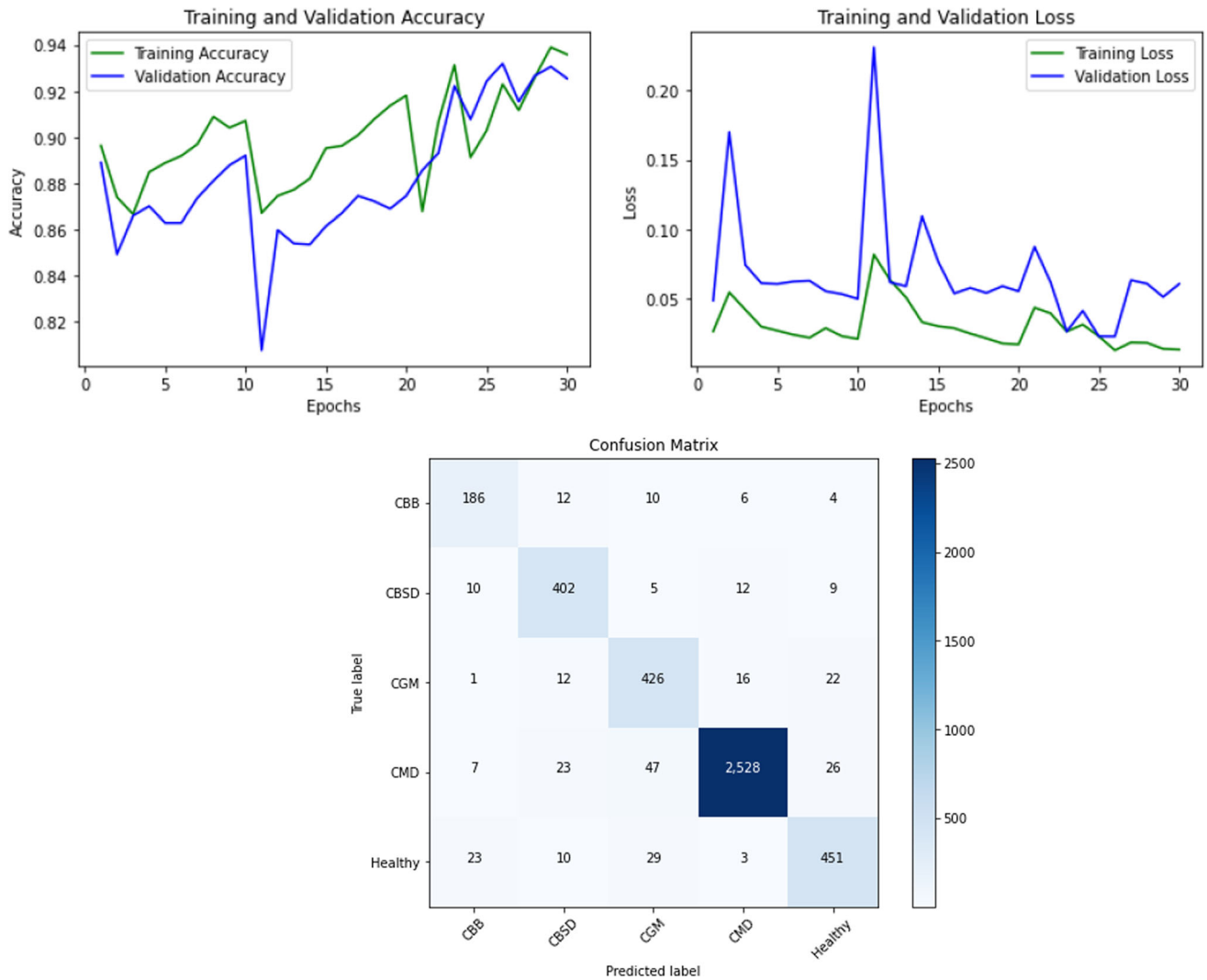


Fig. 9 Analysis of the proposed hybrid dual-path feature fusion network With accuracy and loss graphs. The model was trained for 30 epochs, and it converged well. Moreover, the confusion matrix is also presented in the figure above

learning across all the channels in the network. Furthermore, the depthwise separable convolution focuses on learning the contextual information from the input images and thus improves information flow in the network [6]. As this is a multi-path network, feature maps are aggregated across the network which results in robust and effective feature learning capabilities. As illustrated in Fig. 8, the network achieves an accuracy of 89.9% in classifying the five target classes.

6.4.3 Hybrid dual-path feature fusion network

The proposed network was trained and evaluated for 30 epochs. Various performance metrics were used to evaluate the models, including accuracy, precision, recall, and F1-score. The experimental results of the proposed network

are presented in Fig. 9, along with a class-wise confusion matrix.

The authors proposed an efficient dual-path network to detect diseases in cassava leaves. The RCSANet learns important inter-spatial, inter-channel, and contextual features. Additionally, Efficientnet aids RCSANet by focusing more on channel features with the presence of a squeeze-and-excitation (SE) block. The feature maps from both networks were concatenated to improve the representation power of the network and stability.

To reduce the time for the network to reach optimal convergence and increase performance, weights from the RCSANet were used to initialize model training. As a result, the model achieved a validation accuracy of 93.25% in just 30 epochs. Because cosine annealing with warm restarts was used, the model was able to find the minimum point in a stable region quickly and more precisely.

Particularly for the cassava mosaic disease class, the hybrid dual-path feature fusion network achieved accuracy, precision, recall, and F1-score of 96.73%, 96%, 99%, and 97%, respectively.

High sensitivity in disease detection on cassava leaves plays a major role in assessment of the health of the crop and to further predict crop yield. The network has achieved an overall good recall value of 87.52%, which is important because a high recall value is usually the desired outcome. This indicates the model's propensity toward learning effective feature representations. Despite the presence of a severe class imbalance in the cassava disease dataset, the network performed effectively and produced high-quality results.

Finally, the proposed model was trained and validated with and without data augmentation techniques to prove its effectiveness. Data augmentation is necessary to overcome class imbalance and make the model more robust to variations [25]. The results are presented in Table 3.

7 Discussion

In this section, a visual analysis of the features generated by the proposed hybrid dual-path feature fusion network is presented. Furthermore, the proposed study is compared against state-of-the-art CNN architectures and existing studies.

7.1 Visual interpretation of the features of the proposed network

Understanding and interpreting the neural network and its predictions are crucial for development of an explainable AI model. Table 4 presents a visualization of the gradients generated in the penultimate convolutional layer of the proposed network. gradient-weighted class activation mapping (Grad-CAM) is adopted for the generation of heat maps to visualize the salient regions in the cassava leaf images. Here, the channel-wise feature map gradients are averaged based on their weights at a convolutional layer. Usually, CNNs preserve spatial information, which is lost in fully connected layers, and therefore, the gradients generated in the last convolution layer were selected. The heatmap for the CBB and CGM classes shows an intense band of activations in the diseased regions. For the CBSD

and CMD classes, the activations are rather mild and spread out, as illustrated in Table 4.

Table 5 summarizes the experimental results of the proposed research study. The methods were evaluated with reference to a baseline feedforward CNN architecture that comprises five convolutional layers. The resultant observations prove that RCSANet shows notable improvement over the pre-trained Efficientnet architecture in all the metrics. It can also be observed that the baseline feedforward CNN provides satisfactory results due to its poor feature learning capability. Lastly, the proposed hybrid dual-path feature fusion network significantly outperforms all the other architectures.

Quantitative results of hyperparameter tuning of the focal loss parameters, namely alpha (α) and gamma (γ), are presented in Table 6. First, the default values of the hyperparameters were selected ($\alpha = 0.8$, $\gamma = 2$). Based on the results, the hyperparameters were optimized to produce the highest results. Finally, focal loss with $\alpha = 0.5$ and $\gamma = 2.5$ gave the highest results. The cross-entropy loss was also tested to validate its performance against focal loss.

7.2 Comparison with the state-of-the-art networks

Table 7 presents a comparison of the proposed study with the state-of-the-art CNN architectures pre-trained on the imagenet dataset. Extensive experiments were performed to validate the performance improvements of the proposed network over standard architectures. The pre-trained architectures were fine-tuned to adapt to the cassava dataset. The proposed RCSANet and hybrid dual-path feature fusion network have shown significant improvement in all the metrics over the pre-trained architectures.

Of all the compared state-of-the-art architectures, Resnext-50 performed the best, followed by Efficientnet with noisy student training. Resnext is an extension of the residual network where the standard residual block is replaced with an inception-like block to reduce the computational overhead and extract features more effectively [35]. Xception, on the other hand, performed poorly on the cassava disease dataset with low values of F1-score, precision, and recall. Moreover, with relatively fewer parameters, RCSANet managed to improve on Resnext by 2.5% in accuracy and a factor of 4% on all the other metrics. The proposed hybrid dual-path feature fusion network advanced

Table 3 Comparison of the proposed model performance with and without data augmentation

Experiments	Accuracy in (%)	Precision in (%)	Recall in (%)	F1-score in (%)
Without data augmentation	90.66	88.65	84.50	86.45
With data augmentation	93.25	89.88	87.52	88.68

Table 4 Visualization of the diseased regions on the cassava leaf image using Gradient-weighted class activation mapping (Grad-CAM)


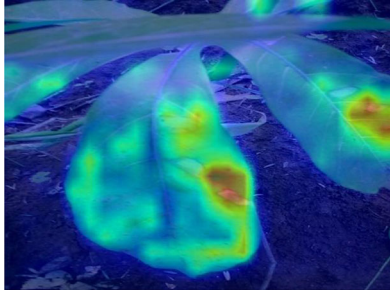



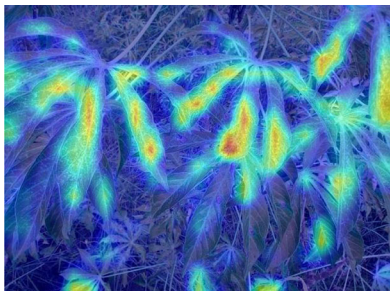
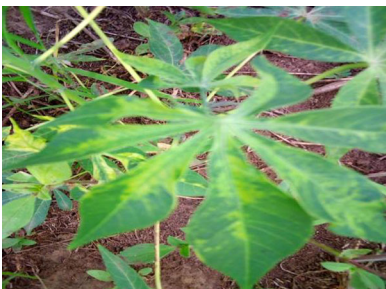

S. No.	Class	Original image	Visualization
1	CBB		
2	CBSD		
3	CGM		
4	CMD		

Table 5 A summary of the ablation studies is presented to validate the effectiveness of the proposed network

Experiments	Accuracy in (%)	Precision in (%)	Recall in (%)	F1-score in (%)
Baseline 5-layer feedforward CNN	63.6	58.5	50.0	54.0
Efficientnet-B0	87.7	81.0	80.0	80.0
RCSANet	89.9	85.0	83.0	84.0
Hybrid dual-path feature fusion network	93.25	89.88	87.52	88.68

Different baseline models were trained on the cassava dataset, that includes a simple 5-layer CNN, Efficientnet, and RCSANet

the results of the RCSANet by a factor of 4.65%. An improvement of 8.68%, 9.88%, and 8.52% over Resnext

can be observed in F1-score, precision, and recall, respectively. On the whole, it could be inferred from

Table 6 Hyperparameter tuning results of focal loss with the proposed network on the cassava disease dataset

Loss	Accuracy in (%)	Precision in (%)	Recall in (%)	F1-score in (%)
Cross-entropy	93.14	87.94	86.50	87.76
Focal loss ($\alpha = 0.8, \gamma = 2$)	89.55	84.35	83.20	84.10
Focal loss ($\alpha = 0.5, \gamma = 2$)	92.54	88.60	87.45	88.56
Focal loss	93.25	89.88	87.52	88.68

Table 7 Quantitative performance comparison of multiple state-of-the-art CNN architectures on the cassava disease dataset using various classification metrics and the number of trainable parameters

Methods	Number of parameters (M)	Macro-average accuracy in (%)	Macro-average F1-score in (%)	Macro-average precision in (%)	Macro-average recall in (%)
Xception	22.8	77.3	58.0	60.0	56.0
VGG16	138	77.7	60.0	62.0	59.0
Alexnet	57	78.2	59.0	60.0	57.0
Densenet121	7	78.9	62.0	65.0	58.0
Resnet50	23	84.4	71.0	74.0	70.0
Efficientnet-B0 (noisy student)	4	84.4	72.0	74.0	70.0
Resnext50	25	87.4	80.0	80.0	79.0
Efficientnet-B0	4	88.2	82.0	81.0	81.0
RCSANet	20	89.9	84.0	85.0	83.0
Hybrid dual-path network	23	93.25	88.68	89.88	87.52

Table 8 Performance analysis of the proposed study with current research studies that have utilized the same Kaggle cassava leaf disease dataset with 21 k samples and the same data distribution

Source	Method	Model	Classes	Number of Cassava samples	Data split	Accuracy in (%)
Maryum et al. [15]	Transfer Learning	Efficientnet-B4	5	21,367	85:15 (Train:Val)	89.09
Chen et al. [5]	Transfer Learning	ResNest-50 (Best)	5	21,367*	fivefold Cross Validation	89.7
Vijayalata et al. [36]	Transfer learning	Efficientnet-B0	5	21,367	80:20 (Train:Val)	92.6
Thai et al. [33]	Transfer learning	Vision transformer	5	21,367	80:20 (Train:Val)	90.0
Methil et al. [19]	Transfer learning	Efficientnet-B4	5	21,367	80:20 (Train:Val)	85.64
Proposed model	Custom CNN	Hybrid dual-path feature fusion network	5	21,367	80:20 (Train:Val)	93.25

Tables 7 and 8 that the hybrid dual-path feature fusion network achieved the best overall performance.

7.3 Performance analysis with the existing studies

The performance of the proposed approach is compared against the existing studies for cassava disease detection.

For a fair comparison, the studies that employed the same cassava dataset were only included. The compared studies have performed classification in five classes (CBB, CBSD, CGM, CMD, and healthy). The evaluation metrics used in this study to compare and analyze the performance with other studies are (a) accuracy, (b) precision, (c) recall, and (d) F1-score.

The existing studies applied transfer learning methods for the detection of cassava plant diseases. Although transfer learning is an effective strategy, different initial and target domains may lead to negative transfer. It is to be noted that the proposed RCSANet is the first custom CNN that is designed and implemented from scratch specifically for this dataset. The classification accuracies obtained by the existing studies are in the range of 89–92.6%. In contrast, the proposed network has shown improved results. Overall, the proposed study outperformed the existing state-of-the-art approaches with an overall accuracy of 93.25%. Furthermore, the feature maps from the two architectures were fused, which results in a network with greater performance and precise feature learning.

8 Conclusion

This research presents a dual-path network to fuse hybrid features from two networks, namely RCSANet and Efficientnet. Most of the existing studies have overlooked the importance of contextual, inter-channel, and inter-spatial features. Therefore, RCSANet is proposed to specifically extract hybrid contextual features from the cassava images. The RCSA module consists of triplet attention that extracts inter-channel and inter-spatial features using cross-dimensional interaction. The proposed RCSANet also consists of auxiliary branches with depthwise separable convolution and average pooling to learn significant relationships between neighboring pixels. To further enhance the feature extraction capability of the network, a feature fusion method is proposed to utilize the strength of both RCSANet and Efficientnet. The proposed hybrid dual-path feature fusion network outperforms the existing methods with an accuracy of 93.25%. This framework can be extended for the diagnosis of infection in other crops such as rice, wheat, and apple. Furthermore, knowledge distillation can also be adopted for the effective detection of plant diseases using smaller architectures. The proposed method can be integrated with drones to identify diseased crops in fields, which efficiently reduces the workload of farmers.

Funding This research did not receive any funding.

Data availability The datasets were sourced from Kaggle. <https://www.kaggle.com/competitions/cassava-leaf-disease-classification>.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

References

1. Abayomi-Alli OO, Damaševičius R, Misra S, Maskeliūnas R (2021) Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning. *Expert Syst.* <https://doi.org/10.1111/exsy.12746>
2. Agricultural Research Council (2014) <https://www.arc.agric.za/arc-iic/Pages/Cassava.aspx>
3. Ayu HR, Surtono A, Apriyanto DK (2021) Deep learning for detection cassava leaf disease. *J Phys Conf Ser* 1751(1):012072. <https://doi.org/10.1088/1742-6596/1751/1/012072>
4. Buslaev A, Iglovikov VI, Khvedchenya E, Parinov A, Druzhinin M, Kalinin AA (2020) Albumentations: fast and flexible image augmentations. *Information* 11(2):125. <https://doi.org/10.3390/info11020125>
5. Chen Y, Xu K, Zhou P, Ban X, He D (2022) Improved cross entropy loss for noisy labels in vision leaf disease classification. *IET Image Process.* 16(6):1511–1519. <https://doi.org/10.1049/ipr2.12402>
6. Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). <https://doi.org/10.1109/cvpr.2017.195>
7. Fanou AA, Zinsou VA, Wydra K (2018) Cassava bacterial blight: a devastating disease of Cassava. In: Cassava. InTech. <https://doi.org/10.5772/intechopen.71527>
8. Ganaie MA, Hu M, Malik AK, Tanveer M, Suganthan PN (2022) Ensemble deep learning: a review. *Eng Appl Artif Intell* 115:105151. <https://doi.org/10.1016/j.engappai.2022.105151>
9. Gao F, Sa J, Wang Z, Zhao Z (2021) Cassava disease detection method based on EfficientNet. In: 2021 7th international conference on systems and informatics (ICSAI). IEEE. <https://doi.org/10.1109/icsai53574.2021.9664101>
10. Hassan SM, Maji AK (2022) Plant disease identification using a novel convolutional neural network. *IEEE Access* 10:5390–5401. <https://doi.org/10.1109/access.2022.3141371>
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr.2016.90>
12. Kaggle competition (2019) <https://www.kaggle.com/competitions/cassava-leaf-disease-classification>
13. Lin M, Chen Q, Yan S (2013) Network in network (Version 3). arXiv <https://doi.org/10.48550/ARXIV.1312.4400>
14. Lozano JC, Booth RH (1974) Diseases of Cassava (*Manihot esculenta* Crantz). *PANS Pest Artic News Summ* 20(1):30–54. <https://doi.org/10.1080/09670877409412334>
15. Maryum A, Akram MU, Salam AA (2021) 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. <https://doi.org/10.1109/honet53078.2021.9615488>

16. Mathulapransan S, Lanthong K (2021) Cassava leaf disease recognition using convolutional neural networks. In: 2021 9th international conference on orange technology (ICOT). IEEE. <https://doi.org/10.1109/icot54518.2021.9680655>
17. McCallum EJ, Anjanappa RB, Gruissem W (2017) Tackling agriculturally relevant diseases in the staple crop cassava (*Manihot esculenta*). *Curr Opin Plant Biol* 38:50–58. <https://doi.org/10.1016/j.pbi.2017.04.008J>
18. Megha M, Chinnapani K, Samala N (2021) Detection of Casava plant related diseases using deep learning. *Int Res J Plant Sci*. <https://doi.org/10.14303/irjps.2021.16>
19. Methil A, Agrawal H, Kaushik V (2021) One-vs-all methodology based Cassava leaf disease detection. In: 2021 12th international conference on computing communication and networking technologies (ICCCNT). IEEE. <https://doi.org/10.1109/icccnt51525.2021.9579920>
20. Metlek S (2021) Disease detection from cassava leaf images with deep learning methods in web environment. *Int J 3D Print Technol Digit Ind*. <https://doi.org/10.46519/ij3dptdi.1029357>
21. Misra D, Nalamada T, Arasanipalai AU, Hou Q (2021) Rotate to attend: convolutional triplet attention module. In: 2021 IEEE winter conference on applications of computer vision (WACV). IEEE. <https://doi.org/10.1109/wacv48630.2021.00318>
22. Mwebaze E, Gebre T, Frome A, Nsumba S, Tusubira J (2019) iCassava 2019 fine-grained visual categorization challenge (Version 2). *arXiv* <https://doi.org/10.48550/ARXIV.1908.02900>
23. Oyewola DO, Dada EG, Misra S, Damaševičius R (2021) Detecting cassava mosaic disease using a deep residual convolutional neural network with distinct block processing. *PeerJ Comput Sci* 7:e352. <https://doi.org/10.7717/peerj-cs.352>
24. Patike KR, Sandeep K, Sreenivasulu K (2021) Cassava leaf disease classification using separable convolutions UNet. *Turk J Comput Math Educ*. <https://doi.org/10.17762/turcomat.v12i7.2554>
25. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning (Version 1). *arXiv* <https://doi.org/10.48550/ARXIV.1712.04621>
26. Ramcharan A, Baranowski K, McCloskey P, Ahmed B, Legg J, Hughes DP (2017) Deep learning for image-based cassava disease detection. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2017.01852>
27. Ramcharan A, McCloskey P, Baranowski K, Mbilinyi N, Mrisho L, Ndalawa M, Legg J, Hughes DP (2019) A mobile-based deep learning model for cassava disease diagnosis. *Front Plant Sci*. <https://doi.org/10.3389/fpls.2019.00272>
28. Ravi V, Acharya V, Pham TD (2021) Attention deep learning-based large-scale learning classifier for Cassava leaf disease classification. *Expert Syst*. <https://doi.org/10.1111/exsy.12862>
29. Sambasivam G, Opiyo GD (2021) A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egypt Inf J* 22(1):27–34. <https://doi.org/10.1016/j.eij.2020.02.007>
30. Sangbamrung I, Praneetpholkrang P, Kanjanawattana S (2020) A novel automatic method for Cassava disease classification using deep learning. *J Adv Inf Technol* 11(4):241–248. <https://doi.org/10.12720/jait.11.4.241-248>
31. Surya R, Gautama E (2020) Cassava leaf disease detection using convolutional neural networks. In: 2020 6th International Conference on Science in Information Technology (ICSITech). IEEE. <https://doi.org/10.1109/icsitech49800.2020.9392051>
32. Tan M, Le QV (2019) EfficientNet: rethinking model scaling for convolutional neural networks. *arXiv* <https://doi.org/10.48550/ARXIV.1905.11946>
33. Thai HT, Tran-Van NY, Le KH (2021) Artificial cognition for early leaf disease detection using vision transformers. In: 2021 international conference on advanced technologies for communications (ATC). IEEE. <https://doi.org/10.1109/atc52653.2021.9598303>
34. Tomlinson KR, Bailey AM, Alicai T, Seal S, Foster GD (2017) Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Mol Plant Pathol* 19(5):1282–1294. <https://doi.org/10.1111/mpp.12613>
35. Xie S, Girshick R, Dollar P, Tu Z, He K (2017) Aggregated residual transformations for deep neural networks. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr.2017.634>
36. Vijayalata Y, Billakanti N, Veeravalli K, Deepa A, Kota L (2022) Early detection of Casava plant leaf diseases using EfficientNet-B0. In: 2022 IEEE Delhi section conference (DELCON). IEEE. <https://doi.org/10.1109/delcon54057.2022.9753210>
37. Yang X (2020) An overview of the attention mechanisms in computer vision. In *J Phys Conf Ser* 1693(1):012173. <https://doi.org/10.1088/1742-6596/1693/1/012173>
38. Zhang X, Zhou X, Lin M, Sun J (2018) ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In: 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr.2018.00716>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.