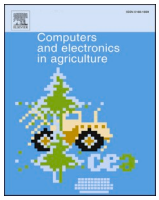
Computers and Electronics in Agriculture 225 (2024) 109351

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



MAIANet: Signal modulation in cassava leaf disease classification

Jiayu Zhang [a], Baohua Zhang [a], Chao Qi [a], Innocent Nyalala [a,d], Peter Mecha [a], Kunjie Chen [a,*], Junfeng Gao [b,c,*]

[a] College of Engineering, Nanjing Agricultural University, Nanjing, 210000, Jiangsu Province, China
[b] Lincoln Agri-Robotics Centre, Lincoln Institute for Agri-Food Technology, University of Lincoln, Lincoln, UK
[c] Lincoln Centre for Autonomous Systems, University of Lincoln, Lincoln, UK
[d] Faculty of Science, Department of Computer Science, Egerton University, Njoro, Kenya

ARTICLE INFO

Keywords:
Disease classification
Multiattention
Feature fitting optimization
Quantization optimization

ABSTRACT

Cassava is the third largest source of carbohydrates for human consumption worldwide; however, it is highly susceptible to viral and bacterial diseases, which pose a significant threat to food security. The advancement of deep learning algorithms in precision agriculture holds the key to enabling the early classification of plant diseases, thereby leading to enhanced crop yields and ultimately stabilizing food security. In the coarse-grained label discrimination task of weak supervision learning, high-quality semantic features contain abundant semantic description information, which plays a crucial role in constructing a precise description of plant disease discrimination in tanglesome field circumstances and directly influences the performance of neural networks. Thus, a multiattention IBN anti-aliasing neural network (MAIANet) was proposed to improve the classification accuracy of cassava leaf disease classification by improving the feature quality in the coarseness label classification task. The proposed MAIANet neural network includes two innovative approaches. First, the multiattention method was designed to scale the feature signals twice to adjust the angular frequency of the feature signals in the residual branch for optimal feature fitting within the residual unit. Second, the anti-aliasing block extracts the high-frequency component feature and optimizes the quantization result of the pooling operation to depress the aliasing signal in the down-sampled feature maps. When the proposed method was tested and validated on the cassava dataset, the results showed that the prediction accuracy of the proposed method significantly improved, with an accuracy of 95.83 %, a loss of 1.720, and an F1-score of 0.9585, outperforming V2-ResNet-101, EfficientNet-B5, RepVGG-B3g4, and AlexNet with significant margins. Based on the above experimental results, the proposed algorithm is suitable for classifying cassava leaf diseases.

1. Introduction

Cassava (*Manihot esculenta* Crantz) is one of the most widely grown crops in tropical and subtropical areas worldwide, and is a major staple food crop that feeds approximately 800 million people worldwide (Chisenga et al., 2019; Howeler et al., 2013) in Africa (53.5 %), Asia (34.7 %), America (11.8 %), and Oceania (0.03 %) (Food and Agriculture Organization of the United Nations, https://faostat.fao.org). Cassava is not only a major food source in the developing world but also an important raw material for producing feed, starch, and ethanol fuel. However, diseases severely affect cassava growth and potentially lead to great yield loss, causing irreversible economic losses for farmers.

Cassava diseases often exhibit similarities and propagate rapidly, leaving farmers to grapple with a diverse array of cassava leaf diseases that encompass over 30 recognized types (Legg et al., 2015). Among these, cassava bacterial blight (CBB), cassava brown streak disease (CBSD), cassava mosaic disease (CMD), and cassava green mottle (CGM) are particularly devastating, contributing significantly to yield losses in cassava crops (McCallum et al., 2017; Ramcharan et al., 2017). Consequently, classification of these four diseases has become a pivotal

Abbreviations: CNN, Convolutional neural network; ms, Millisecond; UAV, Unmanned Aerial Vehicle; LSTM, Long-short term memory; MANet, Multiattention neural network; MABNet, Multiattention batch normalization neural network; MAIBNet, Multiattention instance batch normalization neural network; MAIANet, Multiattention instance batch normalization anti-aliasing neural network; IN, Instance normalization; BN, Batch normalization; IBN, Instance batch normalization; L2-RE, L2-regularization; OS, operation system; 2D, two dimension.

* Corresponding author at: College of Engineering, Nanjing Agricultural University, Nanjing, 210000, Jiangsu Province, China.
E-mail addresses: kunjiechen@njau.edu.cn (K. Chen), Jugao@lincoln.ac.uk (J. Gao).

https://doi.org/10.1016/j.compag.2024.109351
Received 8 January 2024; Received in revised form 20 July 2024; Accepted 11 August 2024
Available online 20 August 2024

research priority. The incidence of cassava mosaic virus epidemics has escalated in East Africa for decades, particularly the brown streak virus (CBSD), resulting in production losses of up to 47 % and annual economic losses of US$ 60 million (in lost yield), leading to local famines. This situation has spurred significant investments in plant breeding programs aimed at addressing this issue (Legg et al., 2015). Cassava bacterial blight disease (CBB) poses a significant challenge to global cassava cultivation, with losses exceeding 50–75 % in regions where highly susceptible cultivars are grown (Wydra and Verdier, 2002). Fig. 1 illustrates examples of cassava disease. The adverse effects of the four cassava diseases examined in this study are shown in Table 1.

After a crop is infected by a disease, its external morphology and internal physiological characteristics undergo significant changes. Symptoms, such as chlorosis, discoloration, deformation, curling, and wilting, become apparent. Internally, physiological adjustments occur in moisture content, pigment levels, photosynthesis, respiration, and defense enzyme systems. By examining these changes in both external and internal characteristics of the crop following disease onset, it is possible to detect and assess the infection status of the crop. Manual classification, initially established as the fundamental method for crop disease classification (Samborski et al., 2009), relies heavily on observations of morphological features of pathogens and accumulated past experience. Although straightforward and intuitive, it is highly subjective and greatly depends on the expertise of the crop specialists. Traditional molecular methods such as polymerase chain reaction (PCR) and enzyme-linked immunosorbent assay (ELISA) have also been developed for crop disease inspection. These classical biosensing techniques are rapid and accurate, but their high cost precludes their widespread adoption by farmers (Martinelli et al., 2015). Subsequently, several emerging imaging methods have been employed for plant disease diagnosis, including visible and near-infrared (Polder et al., 2014), infrared (Zhu et al., 2018) and spectral-based detection methods (Qi et al., 2023). Manual detection fails to meet the efficiency and cost requirements in real-world scenarios. Although biosensing and hyperspectral techniques have proven effective for plant disease diagnosis, their high cost hinders their practical application. In contrast, visible image-based detection techniques offer lower costs and a broader range of applications. Therefore, visible image-based algorithms for plant disease diagnosis have attracted considerable research attention.

Recently, several studies have focused on plant disease classification to facilitate efficient classification solutions for farmers. Various methods have been employed to ensure high-quality semantic feature representation, including image augmentation (Lilhore et al., 2022; Oyewola et al., 2021), feature aggregation (Ravi et al., 2022), and high-level feature extraction (C. R et al., 2022) for classification. Lilhore et al. (2022) introduced depthwise convolution to minimize the feature count and computational load in the head of an enhanced convolutional neural network model. They also utilized a distinct block to process features and address data imbalance, resulting in an outperformance of the enhanced CNN model with an accuracy of 98.7224 %. Oyewola et al. (2021) employed image augmentation to appropriately represent plant disease features in neural networks. Their augmentation solution included methods such as image-distinct block processing and contrast adjustment, leading to improved classification results for cassava disease classification. To enhance the accuracy of plant disease classification further, Ravi et al. (2022) aggregated several high-level features of EfficientNets and employed attention mechanisms to establish long-term dependencies in disease features within a neural network. They also introduced a PCA algorithm to reduce the computing complexity. C. R et al. (2022) utilized the Bi-LSTM algorithm to establish long-term information relationships based on high-level semantic information for cassava leaf disease classification, achieving an F1-score of 0.9549.

Table 1
Harmful effects of the four cassava diseases.

Cassava disease	Hazards to cassava yield
Cassava bacterial blight disease	Yield reduction of 50–75 %. After infection, leaves fall off in large numbers, tubers are small and few, and starch content is reduced (Hillocks and Thresh, 2002; Legg et al., 2015; Wydra and Verdier, 2002).
Cassava mosaic disease	The average yield loss of the infected strains was about 30 %~40 %, and the highest yield loss was 86 % (Hillocks and Thresh, 2002; Legg et al., 2015).
Cassava green mottle disease	A reduction in production of 2 % to 10 % can occur, and in severe cases, it can lead to a reduction of 40 % to 60 % or even cause the entire cassava plant to die (Hillocks and Thresh, 2002; Legg et al., 2015).
Cassava brown streak disease	Production can be reduced by 20 to 25 %, and in severe cases, it can lead to a reduction of up to 70 % (Hillocks and Thresh, 2002; Legg et al., 2015).

The development of high-quality semantic features depends on various neurons within the CNNs (Wei, 2023). Thus, Chang et al. (2020) proposed an innovative loss function to fine-tune neuron parameters, aiming to enhance semantic feature quality, and emphasizing the theory of "devil is in the channels." For crop disease classification task, Zhong and Zhao (2020) integrated DenseNet (Huang et al., 2017) as a backbone with a regression classification algorithm and a multi-classification focal loss function (Lin et al., 2017) to diagnose apple leaf disease, achieving outstanding performance classification performance. The classifier failed to provide a sufficient degree of sensitivity because the number of minority class instances remained small. However, the multiclassification focal loss function cannot completely address the long-tail phenomenon. Thus, Gao et al. (2021) introduced a majority voting method to improve potato late blight lesion segmentation accuracy by accumulating seven masks generated by a SegNet-based encoder-decoder architecture neural network, while also addressing imbalance category distributions of raw images through a weighted loss function. However, the feature quality was not solely determined by the loss function but also depended on the feature forward-propagation architecture. Ma et al. (2018) proposed a symptom-wise classification scheme to enable plant leaf disease classification unaffected by multiple disease symptoms on one leaf. The symptom-wise classification scheme reduced irrelevant leaf feature effects in AlexNet for plant leaf disease

Fig. 1. Illustration of Cassava disease. (Legend: (a) Cassava Bacterial Blight, (b) Cassava Brown Streak disease, (c) Cassava Green Mottle, (d) Cassava Mosaic Disease, (e) Healthy).

classification. However, historical information has not been fully exploited in AlexNet, resulting in its inability to consistently provide high-quality features for the classifier. To explore the relationship between feature quality and plant disease classification performance in complex unstructured environments, Sethy et al. (2020) compared ResNet, MobileNet-V2, and ShuffleNet, finding ResNet to have superior feature quality. They utilized the SVM method for central representation based on ResNet layer feature maps, achieving an F1-score of 0.9838. Although ResNet's residual structure has been classic for historical feature reuse, further evolution of its feature descriptor is necessary. Qi et al. (2023) enhanced the precision of potato late blight disease classification by utilizing a band-screening method to identify the important bands in hyperspectral images. They combined 2DCNN and 3D-CNN based on the concatenation operation of DenseNet and utilized the SE-ResNet method to achieve high-quality lesion features for potato late blight disease classification. In the aforementioned study, high-quality features were considered essential for improving the plant disease classification. However, the relationship between high-quality features and the precision of classification results has not been explored. Therefore, attention was proposed and claimed as an inherent search procedure in visual optimization solutions (Tsotsos et al., 1995).

In early coarse-grained label detection tasks, scholars introduced the concept of selective attention shifting to efficiently extract regions of interest for expressing objective semantic features (Koch and Ullman, 1987). The winner-take-all algorithm plays a crucial role in selective visual attention diversion, filtering saliency regions, and suppressing irrelevant feature expressions. The central expression of the neural network is completed based on the feature information of the salient feature region (Itti et al., 1998). Coarse-grained label detection methods that rely on salient features continue to be explored in the deep learning era (Jerripothula et al., 2016; Tang et al., 2016). In addition to constructing saliency features in feature maps, Zhang et al. (2018) proposed converting top-down attention in CNN classifiers to probabilistic WTA (Winner-Take-All) implementations using the Excitation Backpropagation method. In the realm of plant disease classification using weakly supervised learning, researchers prefer to employ attention algorithms to optimize the quality of the objective semantic features. This enhances the accuracy of convolutional neural networks in classifying plant diseases in the field, facilitating accurate interclass activation mapping and disease-region localization (Chen et al., 2023; Yang et al., 2023). In contrast to weakly supervised neural network algorithms based on attention, Liu et al. (2021b) attempted a clustering weighting-based approach for coarse-grained labeling tasks. They utilized the visual similarity between the same diseases to cluster identical disease patches, weighted the clustering results, and redesigned the loss function for training and validation in the PDD271 plant disease dataset. The loss function significantly affects feature quality in weakly supervised learning tasks for plant disease classification. Li et al. (2019) proposed a weakly supervised learning training strategy based on loss function optimization to address the degradation of neural network performance due to excessive weakly supervised data during training. However, a deeper understanding of deep learning model feature learning techniques or tuning models to learn the focused features is yet to be implemented for crop plant classification (Lee et al., 2017; Rai et al., 2023).

Therefore, to create the focused features in a deep learning model, we devised a neural network rooted in Fourier analysis to craft high-quality semantic features with vein features and other texture features to achieve high-performance cassava leaf disease classification. Our study introduces a multiattention method to modulate feature signals in the identity branch, construct vein features, and integrate them into the identity branch. Additionally, we incorporated an anti-aliasing block to extract high-frequency component features from the vein texture of cassava leaves and applied high-quality pooling features to reduce aliasing signals. This process is followed by downsampling and quantifying the texture features to generate a new feature map. This study utilizes two datasets: the cassava leaf disease dataset and the FGVC-Aircraft dataset. The cassava leaf disease dataset is employed to assess the classification capability of the proposed neural network for cassava leaf disease in field conditions. Additionally, the FGVC-Aircraft dataset is utilized to validate the network's ability in identifying subtle features. The main contributions of this study are threefold.

The semantic features were constructed by observing the veins, outline shapes, and infection features. The vein features were consistently replenished with the feature maps of the identity branch via the residual branch.
The high-frequency component was extracted and used to refine the quantization outcome of the pooling operation when an anti-aliasing block was employed in the proposed neural network.
The irrelevant features were converted into noisy features, which corrupted the semantic features of the leaf region.

The remainder of this paper is organized as follows: Section 2 outlines the dataset and proposed model; Section 3 presents the experimental results; Section 4 offers a discussion and outlines future work; and Section 5 summarizes the findings.

2. Methodology

2.1. Datasets

2.1.1. Cassava disease overview

The incubation period for cassava disease has been proven to be exceptionally long (Zárate-Chaves et al., 2021). Diseases such as Cassava CMD, CBB, CGM, and CBSD can remain latent in cassava seeds (Elango and Lozano, 1980), sand (Maraite, 1993), and other plant hosts (caused by wind- and rain-mediated splashing) (Zárate-Chaves et al., 2021). As the temperature increased, the viruses gradually awakened and spread. Treating major cassava diseases is relatively straightforward for farmers and often involves burning or burying infected material. However, only a small proportion of infected cassava leaves can be treated effectively. The stubborn and challenging nature of cassava leaf diseases necessitates timely classification, discovery, and treatment to minimize losses.

CGM causes white spots on the leaves. It starts with small spots that then enlarge to cover the entire leaf surface, leading to a loss of chlorophyll and affecting photosynthesis. Severe CGM also causes mottling symptoms that can be easily confused with cassava mosaic disease. The affected leaves dry out, shrink, and break away from the plant (Sambasivam and Opiyo, 2021). CBSD symptoms initially include leaf chlorosis, brown streaks on stems, and then dry hard rot in roots, affecting both the quality and quantity of edible storage roots. Its symptoms also include characteristic yellowing of the veins, which sometimes enlarge to form large yellow patches (Hillocks et al., 1996). CMD disease has many foliar symptoms such as mottling, mosaic rust, and twisted leaves, resulting in a general reduction in the sizes of leaves and affected plants (McCallum et al., 2017). Leaves always have patches of green mixed with different colors of yellow and white. These patches reduce the surface area for photosynthesis, resulting in stunted growth and low yield (Abdullahi et al., 2003). CBB symptoms include angular leaf spots, leaf wilting, gum exudates, vascular necrosis of the stem, and then demonstrate shoot dieback. Cassava plants in moist areas are the most affected (McCallum et al., 2017).

2.1.2. Cassava datasets

The cassava leaf disease dataset was collected and annotated by experts from the Uganda National Crops Resources Research Institute (NaCRRI), in collaboration with the AI lab at Makerere University, Kampala (Makerere, 2021). These images were captured by farmers in Uganda using affordable imaging sensors and subsequently released on the Kaggle website on February 19, 2021 (cassava leaf disease dataset, https://www.kaggle.com/competitions/cassava-leaf-disease-classification). The dataset encompasses images with a predominant focus on

intermediate and advanced lesions along with a subset of irrelevant images. Detecting early lesions proved challenging for the low-cost imaging sensors employed, suggesting that spectral sensors may offer greater precision for this task.

The original Uganda cassava leaf disease dataset comprises 21,393 images. However, the initial Uganda cassava dataset lacked balance in its distribution across various categories. The most imbalanced categories, the CMD and CBB disease datasets, contained 13,158 and 1,086 images, respectively. An imbalanced data distribution causes various degrees of long-tail phenomena in feature extraction and class prediction (Chen et al., 2022). To overcome the long-tail obstacle for features and converge the neural network quickly, data augmentation was used to maintain balance in the training dataset to solve data management issues.

Upon analyzing the original Uganda cassava leaf disease dataset, three significant issues require attention. (Azeroual, 2020).

Unmaintained attributes: Unclear and low-quality images of cassava leaf disease make it challenging to clearly distinguish the diseased regions.
Type error: Labeling errors were present in the original cassava leaf disease dataset. It included not only cassava leaves, but also cassava fruits, magazine covers, and other unrelated materials.
Inaccurate data: The lens being out-of-focus results in the loss of high-frequency components in the images. These high-frequency components are crucial for enhancing the generalization of the neural networks. Therefore, inaccurate data poses a significant risk to the success of downstream projects.

Based on the aforementioned issues, there were over 1,000 healthy category images containing disease niduses. Although mislabeled data are common in many benchmark datasets, the unique nature of images within the health category of disease datasets leads to an unacceptable rate of disease-diagnosis errors. The nidus images were removed from the original dataset to ensure the integrity of healthy images.

To maintain balance among the categories, augmentation techniques such as Gaussian noise, horizontal flipping, cutout, and vertical flipping were employed. The 20,000 color images were randomly combined into five balanced categories, and the CMD category was randomly selected from 13,158 images in the raw data. Notably, CMD data were selected without augmentation. Following preprocessing, the images had a resolution of 448 × 448, as outlined in Table 2.

2.1.3. Subtle feature identification dataset

Detecting cassava leaf diseases requires not only weakly supervised classification to identify diseases in complex unstructured environments but also the discernment of subtle differences in features among similar samples (Liu et al. 2021b). The term 'coarse-grained label sample data' refers to annotated samples in which the labels only specify the category of the data, without indicating the location of the target within these samples. Fine-grained image classification presents a challenging problem because of the significant intraclass differences and minimal interclass variance (Wang et al., 2019). Given the high similarity and subtle differences in the characteristics of certain cassava leaf diseases under field conditions, coupled with the considerable variation in disease manifestations at different infection stages, it is challenging for neural network models to accurately differentiate between various crop disease categories. Consequently, cassava leaf disease classification is considered to be a fine-grained feature recognition task that utilizes coarse-grained label samples (Liu et al., 2021b). Therefore, to evaluate the ability of the proposed method to discern subtle feature differences among similar samples, the proposed neural network was validated using the FGVC-Aircraft dataset. This approach also demonstrated the method's superior performance when applied to classifying cassava leaf diseases.

In fine-grained research, the FGVC-Aircraft dataset (Maji et al., 2013) is a well-established benchmark for fine-grained visual classifications. The FGVC-Aircraft dataset was cited in over 1000 papers and utilized as a benchmark dataset in over 200 papers (Paper W, 2024). The dataset contained 10,000 images in three data structures: aircraft manufacturer (30 categories), aircraft family (70 categories), and aircraft variant (102 categories). The aircraft manufacturer format, with its moderate number of categories, allows for the direct training of convolutional neural network models using image augmentation methods. This format includes 3,333 test images and 6,667 training-validation images. After augmentation, the training-validation set for the manufacturer format included 44,010 images across 30 categories, each with 1,467 images, whereas the test dataset remained unaugmented. The image resolution used for the training and testing was 448 × 448 × 3.

Table 2
Dataset analysis.

Augmentation	Base	Noise	Horizontal	Vertical	Cutout	Total
CBB	986	985	986	986	57	4000
CBSD	1772	0	1772	456	0	4000
CGM	1861	28	1861	250	0	4000
CMD	4000	0	0	0	0	4000
Healthy	1054	838	1054	1054	0	4000

2.2. Maianet

In agriculture, navigating unpredictable field conditions and ensuring precise predictions despite obstacles are paramount. This requires neural networks that generalize well and remain robust. Developing a cost-effective algorithm for field cassava leaf disease classification is essential to strike a balance between affordability and performance in agricultural settings. The residual structure of the ResNet model employs residual branches to generate historical feature information (Jin et al., 2000; Weston et al., 2015), and identity branches to capture high-quality semantic features. Compared with other classical neural network models, ResNet has a simple and efficient structure with exceptional mechanisms for transmitting and reusing historical feature information. ResNet maintains a balance between classification accuracy, inference speed, training speed, and parameter count, showing stable performance and good scalability in previous research (BS, 2022; Oyewola et al., 2021). Considering the temporal and spatial complexities of neural networks, ResNet-101 was the most suitable choice in this study, outperforming ResNet-152 and other ResNet-based structures (Gao et al., 2019; Wu et al., 2019).

The architecture of the MAIANet-3 neural network, shown in Fig. 2, was implemented and expanded on the V2-ResNet-101 architecture. The downsampling of ResNet-101 was modified to utilize convolution with a stride of two (NVIDIA, 2023), and the fully connected layer was replaced with a global average pooling layer, resulting in the creation of V2-ResNet-101 for this study. The architecture of MAIANet-3 is detailed in Table 14 in Appendix E. Utilized for cassava disease classification. The MAIANet-3 101-layer neural network is denoted as MAIANet-3 in this study. It comprises three basic components: a head block, anti-aliasing block, and MAIA block.

For a detailed description of the head-block architecture, please refer to Section 2.2.1. For a detailed description of the MAIA block architecture, please refer to Section 2.2.2. For a detailed description of the anti-aliasing block architecture, please refer to Section 2.2.3.

2.2.1. Head block

The head block proposed by He et al. (2016) was employed to extract low-level features from the input image and to downsample the input image twice. This block comprised one 2-D pooling convolution layer, one batch normalization layer, one ReLU activation layer, and one 2-D max-pooling layer.

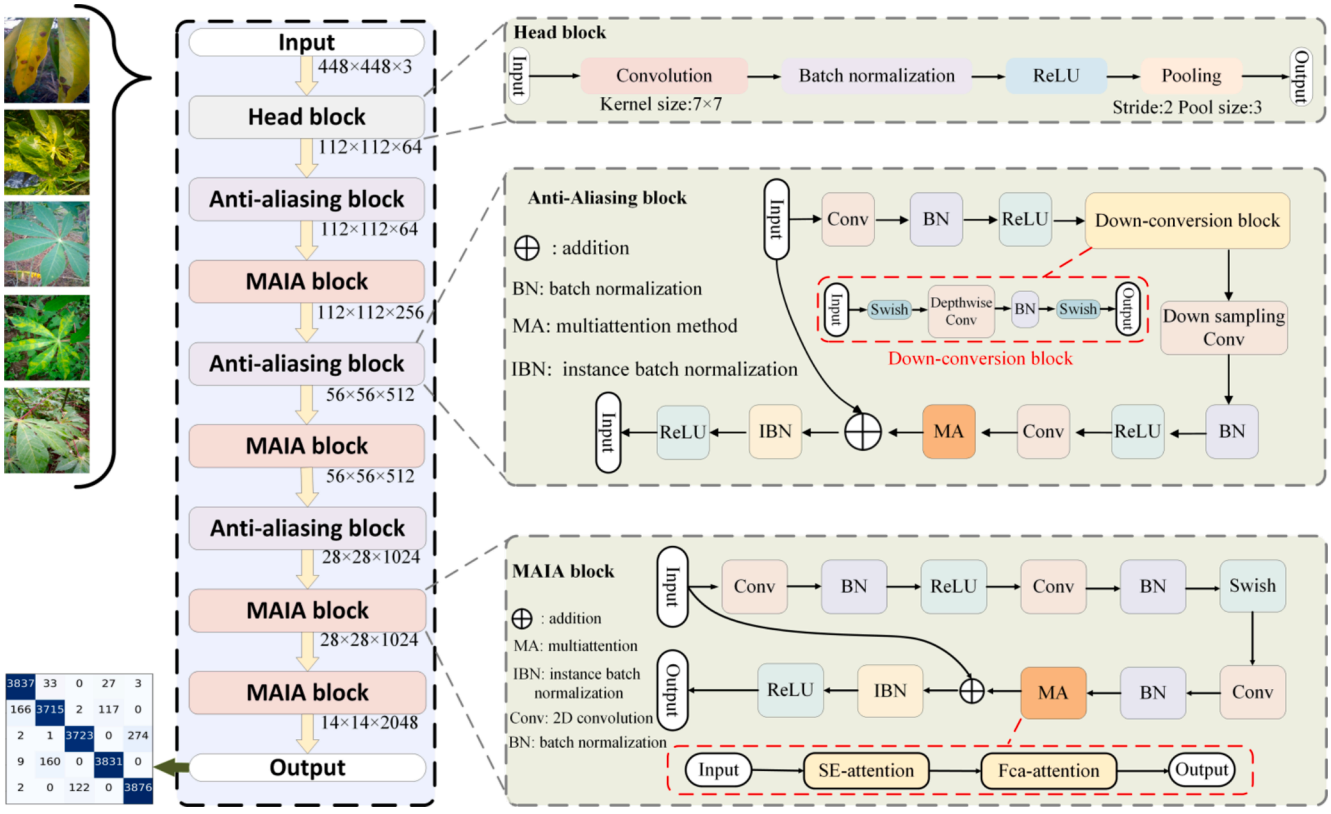


Fig. 2. Architecture for implementing of MAIANet-3.

2.2.2. *MAIA block*

The MAIA block introduces two innovative elements to enhance the feature representation. First, the multiattention block (MA block) was proposed and utilized to adjust the angular frequency of the feature maps, thereby crafting high-quality semantic features with veins and other texture features. These features were stacked in the identity branch to construct a complete semantic feature in the feature maps. Second, instance batch normalization was utilized to adjust the distribution of the feature values for the tensor addition output feature maps of the MAIA block. The architecture of the MAIA block is illustrated in Fig. 2.

2.2.2.1. Multiattention method. Over the past decade, channel attention algorithms have been widely used in neural networks. However, the literature reviewed did not address the issue of fitting bias associated with the scale coefficient in channel attention. In addition, the factors influencing the scale coefficient of channel attention have not been thoroughly explored. To explore the effect of the scale coefficient in the channel attention method, we introduced the optimal approximation theorem of projection distance (Appendix F) and established a mathematical expression of the channel scale (Appendix A) in the Fourier analysis. The theorem of the optimal approximation theorem of projection distance revealed that the number of Fourier coefficients and the value of the Fourier coefficient are essential impact factors for feature fitting. By leveraging the aforementioned theorems, we propose a scale coefficient correction method called the multiattention method.

The mathematical expression of multiattention is presented in Section 2.2.2.1.1. The modulated feature signals were utilized to stack the identity branch of the MAIA blocks in the MAIANet-3 stage. Thus, the mathematical expression of the Fourier coefficient revision for feature signal stacking is presented in Section 2.2.2.1.2. By adjusting both the angle frequency and Fourier coefficient within the residual unit, the quality of the feature signal is enhanced.

2.2.2.1.1. Rectifying scale coefficient in multiattention method. The multiattention method (MA) was constructed using a SE-block (Hu et al., 2018) and a Fca-block (Qin et al., 2021). A multiattention block is a computational unit that can be built on a transformation to map an input tensor $X \in \mathbb{R}^{H\times W\times C}$ to $U \in \mathbb{R}^{H\times W\times C}$. The SE-block outputs can be written as, $U = [u_1, \cdots, u_c]$ the Fca-block outputs as $K = [k_1, \cdots, k_c]$ and the multiattention computing equations as shown in Equations (1) and (2).

$$u_c = X^* v_c = \sum_{s=1}^{C'} X^s * v_c^s \tag{1}$$

$$k_c = U^* v_c = \sum_{s=1}^{C'} U^s * \tau_c^s \tag{2}$$

where * denotes the convolution, $v_c = [v_c^1, \cdots, v_c^{C'}]$, $X = [x^1, \cdots, x^C]$ and $U = [u^1, \cdots, u^C]$. τ_c^s and v_c^s are the 2-D (2 dimensions) convolution kernel in the kernel list v_c and τ_c that convoluted on the corresponding channel of X and U. Bias terms were omitted to simplify the notation. The abovementioned equation can be rewritten as the following mathematical expression for Equation (3):

$$k_c = \sum_{s=1}^{C'} X^s * v_c^s * \tau_c^s = \sum_{s=1}^{C'} X^s * \rho_c^s \tag{3}$$

where, X^s refers to the input tensor, k_c refers to the output tensor, and C' refers to the account of the feature maps in the input tensor. The ρ_c^s is the 2-D (2 dimensions) convolution kernel in the kernel list ρ_c.

Because a two-dimensional signal is composed of frequency-domain K-space signals from two sets of one-dimensional signals in different directions, it becomes difficult to directly calculate the Fourier coefficients and angular frequencies of the two-dimensional feature signals (Moratal et al., 2008). Thus, modulation should be based on one-dimensional feature signals during the frequency modulation process of two-dimensional feature signals (Katz and Bar-Ness, 2015; Wang et al., 2002). Based on the Fourier analysis derivation of the channel

scale in the channel attention method of Appendix A, the mathematical expression of multiattention can be rewritten as Equation (4).

$$\begin{aligned}\delta(t) &= \bar{g}(t)^{*}\rho(t-T_0)\\ &= \rho\times\frac{1}{T}\sum_{n=-\infty}^{\infty} g(jn\omega_0)e^{jn\omega_0 t},\ \omega_0=\frac{2\pi}{T}\end{aligned} \tag{4}$$

where g refers to the one-dimensional feature signal, $\bar{g}$ refers to the result of periodic extension, ρ refers to the scale coefficient, n refers to the scale number, ω_0 refers to the angle frequency, T refers to the period of the feature signal, and t refers to the variable of one-dimensional feature signal.

Based on the mathematical derivation in Appendix A, the angle frequency modulation in the channel scale operation can be further written as Equation (5).

$$\begin{aligned}\delta(t) &= \frac{\rho}{T}\sum_{n=-\infty}^{\infty} g(jn\omega_0)e^{jn\omega_0 t},\ \omega_0=\frac{2\pi}{T}\\ &= \frac{1}{T_\alpha}\sum_{n=-\infty}^{\infty} g(jn\omega_0)e^{jn\omega_0 t},\ \omega_g=\frac{2\pi}{T_\alpha}\ and\ \omega_g\neq\omega_0\end{aligned} \tag{5}$$

where T_α refers to the modulated signal period and ω_g refers to the modulated angle frequency of the signal.

Building on the derivation outlined above, the MA method corrects the scale coefficient of single channel-attention to precisely modulate the angle frequency of the feature signal. As outlined in the projection distance of the Fourier analysis (Appendix F), the projection distance was obtained by adjusting the angular frequency and Fourier coefficient of the fitted signal. Thus, the adjustment of the Fourier coefficient is elaborated in Section 2.2.2.1.2.

2.2.2.1.2. Stacking feature signals for the truth feature signals. As stated by He et al. (2016), the residual unit is given by Equation (6).

$$Y_l = h(x_l) + F(X_l, W_l) \tag{6}$$

where $F(X_l, W_l)$ refers to the residual branch of the residual unit, which is composed of three convolutions, $h(x_l)$ refers to the identity branch of the residual unit, and Y_l refers to the output of the residual unit.

Based on Equation (6), the feature aggregation in the same stage as MAIANet-3 can be rewritten as Equation (7).

$$F_{out} = h^{l}_{Iden} + h^{l}_{Resi} + h^{l+1}_{Resi} + \cdots + h^{l+n-1}_{Resi} \tag{7}$$

where F_{out} represents the output feature maps of the same stage in a neural network, l refers to the first residual unit, h_{Iden} denotes the feature maps in the identity branch, h_{Resi} refers to the feature maps in the residual branch, and n signifies the number of residual units in the same stage of a convolutional neural network. Signal addition is represented by Equation (8).

$$F_{output} = F_{identity} + F_{residual} \tag{8}$$

where F_{output} denotes to the Fourier coefficient of the output feature maps of the residual unit, $F_{identity}$ represents the Fourier coefficient of feature maps in the identity branch, and $F_{residual}$ signifies the Fourier coefficient of feature maps in the residual branch. The definition of Fourier coefficient was provided in Appendix D.

Based on the Fourier coefficient summation theorem of Equation (8), the last $f(Y_l)$ stage can be expressed as Equation (9). A schematic of the key frequency component feature stacking is shown in Fig. 3.

$$F_{out}(jn\omega) = F^{l}_{Iden}(jn\omega) + F^{l}_{Resi}(jn\omega) + F^{l+1}_{Resi}(jn\omega) + \cdots + F^{l+n-1}_{Resi}(jn\omega) \tag{9}$$

where $F^{l}_{Iden}(jn\omega)$ is the Fourier coefficient of the feature signals of the identity branch, $F^{l}_{Resi}(jn\omega)$ is the Fourier coefficient of the feature signals of the residual branch, l is the $l-$th unit, and n is the number of residual units in the same stage. Thus, the semantic feature was fitted to Equation (9). The visualization results of the residual unit are presented in Section 3.2.1.

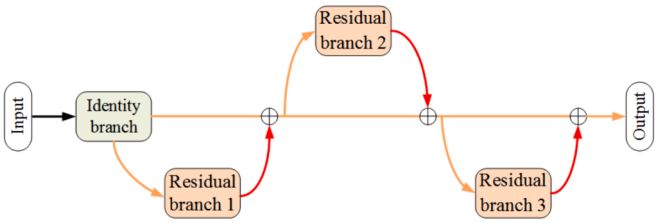


Fig. 3. Key frequency component feature stacking in Stage 1 of MAIANet-3.

2.2.2.2. Normalization features numerical distribution. An efficient normalization method was used to highlight semantic information in the feature maps for high-quality feature representation. High-quality feature representation improves fine-grained discrimination results (Jia et al., 2019; Liu et al., 2021a; Wang et al., 2023). Generally, smoothing calculations (Santurkar et al., 2018) should be effective, simple, and stable to reduce the digital signal mutations. In this regard, instance batch normalization (IBN) (Pan et al., 2018), which is composed of instance normalization (Ulyanov et al., 2016) and batch normalization (Ioffe and Szegedy, 2015), has shown excellent performance with respect to feature divergence variation limits. In this study, the instance batch normalization (IBN) method, which integrates both instance normalization and batch normalization methods, was utilized to enhance the performance of a neural network.

As indicated by Wang et al. (2020), the batch normalization method may not adequately handle the low-frequency components of the convolutional neural network. Consequently, after implementing instance normalization in instance batch normalization, the instance normalization method focused on processing the low-frequency component feature of the neural network, while batch normalization was employed to handle the high-frequency component feature.

The experiment demonstrated a notable increase in accuracy from 90.12 % to 94.99 % with the incorporation of IBN into our algorithm for cassava leaf disease classification, as shown in Table 8. The instance batch normalization method significantly enhances the fine-grained discrimination capabilities of the proposed neural network.

2.2.3. Limiting aliasing signal in neural network

Aliasing refers to the phenomenon in which high-frequency signals are degraded to completely different signals after sampling (Zou et al., 2023). As stated in the *Nyquist sampling theorem* (Nixon and Aguado, 2012), the *Nyquist* sampling frequency should be greater than twice the highest frequency of the signal for the sampled signal to be completely reconstructed into the original signal. However, in convolutional neural networks, the sampling frequency does not always exceed twice the feature signal frequency (Zou et al., 2023). Therefore, optimizing the quantization method is an effective way to enhance the quality of the pooling results.

To transform discrete information from sampled data into continuous information, a quantization method was employed to establish a continuous semantic representation in the feature maps. Optimization of the quantization method involves two distinct steps: semantic feature deconstruction and semantic feature reconstruction in an anti-aliasing block. Deconstruction of the semantic features was processed in the down-conversion block, as described in Section 2.2.3.1. The reconstruction of the semantic feature occurs in the 2D downsampling convolution, as detailed in Section 2.2.3.2.

It was demonstrated that the high-frequency component features boosted the performance of the convolutional neural network in the coarse-grained label classification task. The high-frequency component feature was extracted and represented as the vein feature of the cassava leaf in the down-conversion block and was utilized to fit the pooling feature in the pooling convolution layer.

2.2.3.1. Semantic feature deconstruction in the down-conversion block. As

depicted in Fig. 2, down-conversion was constructed with two swish activation layers (Ramachandran et al., 2017), one batch normalization layer, and one depthwise convolution layer. Depthwise convolution (Chollet, 2017) played a crucial role in the down-conversion block for extracting high-frequency component features, which were represented as vein texture information and other texture information of the cassava leaf. The kernel function of the depthwise convolution matched the specific angular frequency feature signals from the input feature map, and the mathematical derivation was shown in Appendix B.

According to Wang et al. (2020), high-frequency components do not align with human visual preferences. In the evolution of neural networks, abundant low-frequency component data often drives their development. This has led to most current vision neural networks first extracting and learning from the low-frequency component and then learning the high-frequency component to further enhance accuracy. Additionally, as stated by Wang et al. (2020), the high-frequency component is particularly beneficial for small-object discrimination. The visualization results in Fig. 17 of section 3.4.1.1 of down-conversion confirmed this perspective, demonstrating that the high-frequency component feature could improve the accuracy of cassava leaf disease classification, even in coarse-grained label classification tasks.

2.2.3.2. Semantic feature reconstruction in the down-sampling operation. The high-frequency component feature was weighted and aggregated in the 2-D pooling convolution layer, as shown in Equation (10).

$$F_{single\_featuer\_map} = \sum_{0}^{n} [\![(\{f\}_0^n * \{g\}_0^n) * \{h\}_0^n]\!]_{sample} = \sum_{0}^{n} [\![\{f\}_0^n * (\{g\}_0^n * \{h\}_0^n)]\!]_{sample} \quad (10)$$

where $F_{single\_featuer\_map}$ represents the single-feature maps of the 2D downsampling convolution; $\{f\}_0^n$ denotes the feature signal list; $\{g\}_0^n$ refers to the kernel list of the depthwise convolution; $\{h\}_0^n$ represents the kernel list of the 2D downsampling convolution; and n indicates the number of elements in the accumulation operation. The activation function and normalization method were integrated as part of the $\{g\}_0^n$ function. These kernel lists $\{g\}_0^n$ and $\{h\}_0^n$ were utilized to generate a new kernel function list for data weighting, summation, and averaging, and the fitted data were used as the data for the sampled feature maps. The sampled data can be considered as a fitting result based on the local context information. Data that were not fitted were not included. The summation of feature maps also constitutes a fundamental fitting approach.

Based on the mathematical derivation, the pooling kernel function was modified using the kernel function of depthwise convolution. This implies that the high-frequency component feature directly optimizes the pooling outcome. The quality of the quantization result was enhanced by optimizing the kernel function in the pooling operator. Clear semantic information was displayed in the feature maps of the 2D downsampling convolution operator. The visualization results of the downsampling operation are presented in Fig. 18 in Section 3.4.1.2.

2.3. Model training

The proposed neural network was trained on cassava leaf disease datasets using a seven-fold cross-validation method and compared with MAIANet-3, EfficientNet-B5 (Tan and Le, 2019), RepVGG-B3g4 (Ding et al., 2021), V2-ResNet-101 (NVIDIA, 2023), and AlexNet (Krizhevsky et al., 2012).

EfficientNet demonstrated excellent performance on the ImageNet-1 K dataset in 2019 and was established as a universal convolutional neural network. As stated in Ferentinos's (2018) study, the VGG neural network and AlexNet accuracy ranked first and second, respectively, compared to other neural networks. The classical neural network, VGG, was modified into a new structure called RepVGG.

The end-to-end architecture and other algorithms proposed in this study were implemented using TensorFlow 2.4.1 (Abadi et al., 2016) and were run on an AMD Ryzen 7 3800XT CPU @3.89 GHz with a NVIDIA GeForce RTX 3090 GPU. The experimental environment of this study was based on the Windows 10 OS, CUDA 11.2, and cuDNN 8.1.1.

2.3.1. Training with cassava disease dataset

The proposed method was trained on the cassava dataset using the following settings: an SGD optimizer (Zhou et al., 2020) with an initial learning rate of 0.2, decay of 0.96 every epoch, momentum of 0.9, batch normalization momentum of 0.9, and weight decay of 1e-5. L2 regularization (Ng, 2004) was employed as the optimization method in the descriptor of the proposed neural network, with a coefficient set to 1e-5 in the identity branch and 1e-4 in the residual branch. The batch size was set to 12 for MAIANet's neural network training and validation and the training epoch was set to 80. The sigmoid activation function was replaced with a hard-sigmoid activation function in the SE block to reduce the computational cost of the neural network. The proposed method uses categorical cross-entropy as the loss function. Model validation typically involves selecting 70 % of the sample data in the training dataset. In the remaining 30 % of the dataset sample data, 15 % and 15 % were selected as the validation and test data, respectively (Wu et al., 2018). Therefore, 7-fold cross-validation was adopted, with 14.28 % of the dataset samples reserved for data verification for each fold cross-validation.

The TensorFlow framework did not provide an API (Application Programming Interface) for K-fold cross-validation to free up the GPU memory space for the trained model. Therefore, multi-processing is necessary. The process function was used as the enumerator to control the iterations of the 7-fold cross-validation.

2.3.2. Training with the FGVC-Aircraft dataset

The proposed method was trained on the FGVC-Aircraft dataset using the following settings: an SGDW optimizer (Loshchilov and Hutter, 2019) was employed with an initial learning rate of 0.1, weight decay of 1e-5, and momentum of 0.9. The categorical cross-entropy loss function was used with a one-hot code format and label smoothing coefficient of 0.1. The batch size was set to 24 for the MAIANet-3 neural network (based on the MAIANet3-50 layers, detailed in Table 14 of Appendix E) for training and testing. The number of training epochs was set as 420. A piecewise constant decay learning rate schedule was utilized in the fit function of the TensorFlow Keras API (Abadi et al., 2016) (detailed in Table 3). L2-regularization was employed as the optimization method in the descriptor of the proposed neural network, with a coefficient set to 1e-5 in the identity branch and 1e-4 in the residual branch.

2.4. Model evaluation

The performance indicators of the neural networks were computed using four indicators: true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The formulas for the accuracy, recall, precision, and F1-score are given in Equations (11), 12, 13, and 14, respectively:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (14)$$

Table 3
Piece-wise constant decay learning rate table.

Epoch	0 ~ 39	40 ~ 79	80 ~ 149	150 ~ 199	200 ~ 229	230 ~ 249	250 ~ 279	280 ~ 299	300 ~ 349	350 ~ 379	380 ~ 420
Learn rate	0.1	0.06	0.01	0.006	0.004	0.002	0.001	0.0006	0.0004	0.00008	0.00004

3. Results

3.1. Comparison against classical algorithms

Several algorithms were compared in this section, including MAIANet-3, V2-ResNet-101 (NVIDIA, 2023), EfficientNet-B5 (Tan and Le, 2019), AlexNet (Krizhevsky et al., 2012), and RepVGG-B3g4 (Ding et al., 2021). Notably, this comparison did not involve pretraining or ensemble learning. Furthermore, the proposed algorithm was applied to the FGVC-Aircraft benchmark dataset (Maji et al., 2013). The results of the comparison between the algorithms are presented in Table 4.

The loss function (Fig. 4) of the proposed algorithm exhibited a fast gradient descent rate from epochs 1 to 10 and gradually entered a smooth descent state, in which a minimum value of 1.720 was observed. The L2-regularization optimization method caused a high loss value in the initial training stage, and the limited loss value could not be reduced to a very small value. A comparison of the loss curves of the MAIANet-3 neural network without L2-regularization is shown in Fig. 5, and the data analysis is shown in Table 5. In Appendix G, Figs. 22 and 23 (detailed view of Fig. 22) reveal that the accuracy curves of MAIANet-3 without L2 regularization, after fitting to the cassava leaf disease dataset, do not show a clear relationship between the number of training epochs and accuracy improvement. Although MAIANet-3 with L2 regularization shows a gradual increase in accuracy on the cassava leaf disease dataset with more training epochs, the 7-fold cross-validation results reveal only a 0.18 % accuracy improvement compared to MAIANet-3 without L2 regularization. This indicates that L2 regularization slightly improves MAIANet-3′s performance. Additionally, as indicated by the curve trend, with an increase in the number of continuous training epochs, the loss curve continues to decrease, albeit with a very small margin. Considering the balance between training cost and accuracy, we chose to use 80 epochs.

Annotation: the "w/" referred to the utilization of L2-regularization as the optimization method in the MAIANet-3 neural network. In the identity branch, the L2-regularization coefficient was set to 1e-5, and in the residual branch, the L2-regularization coefficient was set to 1e-4. The "w/o" referred to the absence of L2-regularization being utilized in the MAIANet-3 neural network.

The accuracy (Fig. 6), recall (Fig. 7), precision (Fig. 8), and F1-score (Fig. 9) curves of MAIANet-3 are extremely similar. The proposed neural network converged rapidly from epochs 1 to 10, and then converged smoothly until 20 epochs. After 20 epochs, the performance of the neural network slowly improved and the vibration of the indicator curve was reduced. The indicator curve vibration of MAIANet-3 outperformed that of the compared neural networks, particularly after 20 to 80 epochs.

Table 4
Ours method vs. other methods on the cassava dataset using 7-fold cross validation.

Method	Accuracy	Recall	Precision	Loss	F1-score
MAIANet-3	95.83 %	95.81 %	95.82 %	1.720	0.9585
V2-ResNet-101 (NVIDIA, 2023)	77.84 %	77.61 %	78.17 %	1.387	0.7789
EfficientNet-B5 (Tan and Le, 2019)	92.43 %	92.37 %	92.46 %	1.469	0.9242
AlexNet (Krizhevsky et al., 2012)	62.46 %	61.98 %	62.94 %	2.718	0.6246
RepVgg-B3g4 (Ding et al., 2021)	93.08 %	93.07 %	93.14 %	0.347	0.9311

The accuracy of MAIANet-3 is only 2.75 % higher than that of RepVGG-B3g4.

3.1.1. Confusion matrix of MAIANet-3

By aggregating the final validation results of each fold of cross-validation from MAIANet-3, a confusion matrix based on 7-fold cross-validation was obtained. The confusion matrix for the 7-fold cross-validation of the MAIANet-3 neural-network model is shown in Fig. 10. The observation in Fig. 10 reveals that the MAIANet-3 neural network model achieved a high classification accuracy for all types of cassava leaf sample data. Furthermore, in the confusion matrix generated from 7-fold cross-validation of the five classes of cassava leaf sample data, there was no significant fluctuation in accuracy between classes.

3.2. Multiattention block impacts

In an attention ablation experiment, single-channel attention proved ineffective as an attention method for classifying cassava leaf disease. However, the multiattention block demonstrated exceptional performance in this task. The multiattention neural network (MANet) employed the MA block in V2-ResNet-101 and achieved a significant improvement in accuracy, reaching 90.12 %, representing a 3.22 % increase compared to V2-ResNet-101. MANet was compared with V2-ResNet, SENet, FcaNet, and the double SE-block neural network (Double-SENet), with V2-ResNet-101 serving as the baseline neural network in this ablation experiment. The comparison results are presented in Table 7, where MANet achieved the best performance among the compared neural networks.

Considering the performance of V2-ResNet-101 with different random states in the StratifiedKFold function, V2-ResNet-101 with random state 0 achieved a higher accuracy of 9.06 % compared to random state 834,239. The experimental results are presented in Table 6. Random state 834,239 was randomly selected to ensure that the random parameters did not exhibit regular patterns, thereby preventing any potential impact on the data extraction results. Based on the performance comparison in Table 6, the 7-fold schedule in this section sets the random state to zero in the Stratified KFold function of Scikit-Learn. While StratifiedKFold, with a setting of random state 0, performed stratified data sampling, the data path list shuffle was not conducted before data path list sampling. The data were shuffled into the generator function of the Keras framework (Abadi et al., 2016).

It was clearly found that MANet achieved the best performance in the comparison result. However, FcaNet and SENet did not demonstrate any improvement in the generalization of the cassava disease classification. The inclusion of Double-SENet and MANet rectified the scale coefficient twice in the residual branch, resulting in an improved generalization of the neural network. In Section 3.2.1, visualization feature maps of the MA block are provided.

3.2.1. Visualization of the multiattention method in the MAIANet-3

3.2.1.1. Visualization of the residual unit of MAIA block. According to a study conducted by Lee et al. (2017), vein features are the most representative features for plant classification compared with contour features. However, vein features were not the most representative in the coarse-grained label classification task. Nevertheless, this study observed that the vein, outline shape, and infection features constructed the semantic feature, with the vein feature consistently replenished to the feature maps of the identity branch by the residual branch. This

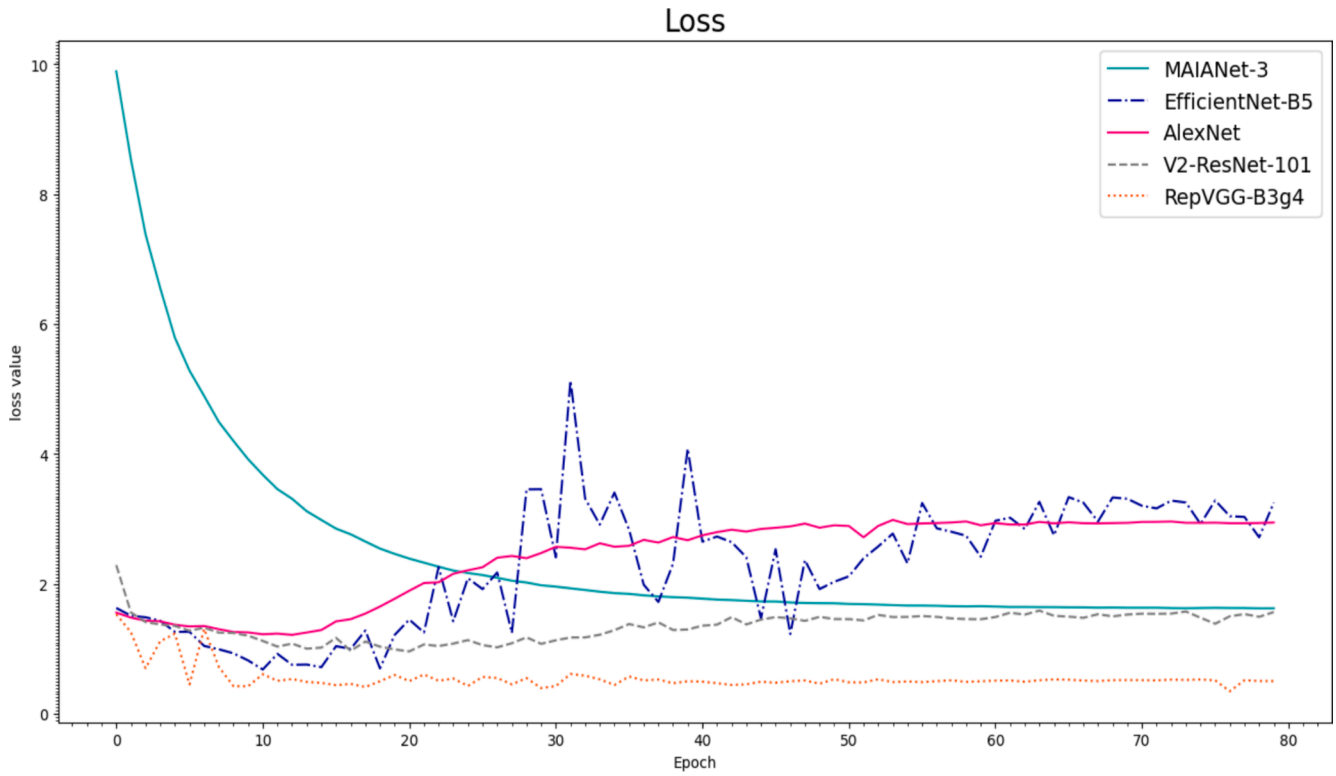


Fig. 4. Validation loss of the algorithms on the cassava dataset using 7-fold cross validation

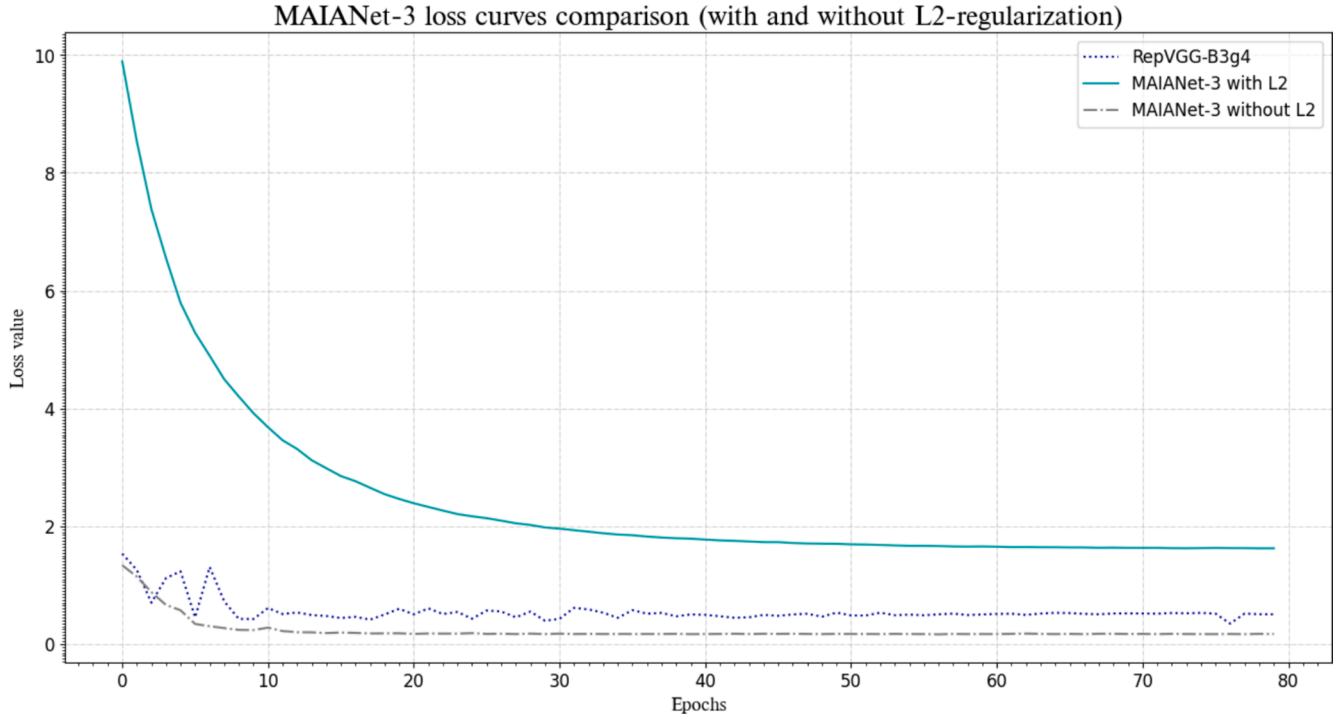


Fig. 5. MAIANet-3 loss comparison on the cassava dataset using 7-fold cross validation

Table 5
Performance comparison with L2-regularization in MAIANet-3 neural network.

Method	Accuracy	Recall	Precision	Loss	F1-score	L2-regularization
MAIANet-3 (w/ L2)	95.83 %	95.81 %	95.82 %	1.720	0.9585	w/
MAIANet-3 (w/o L2)	95.65 %	95.58 %	95.67 %	0.166	0.9562	w/o

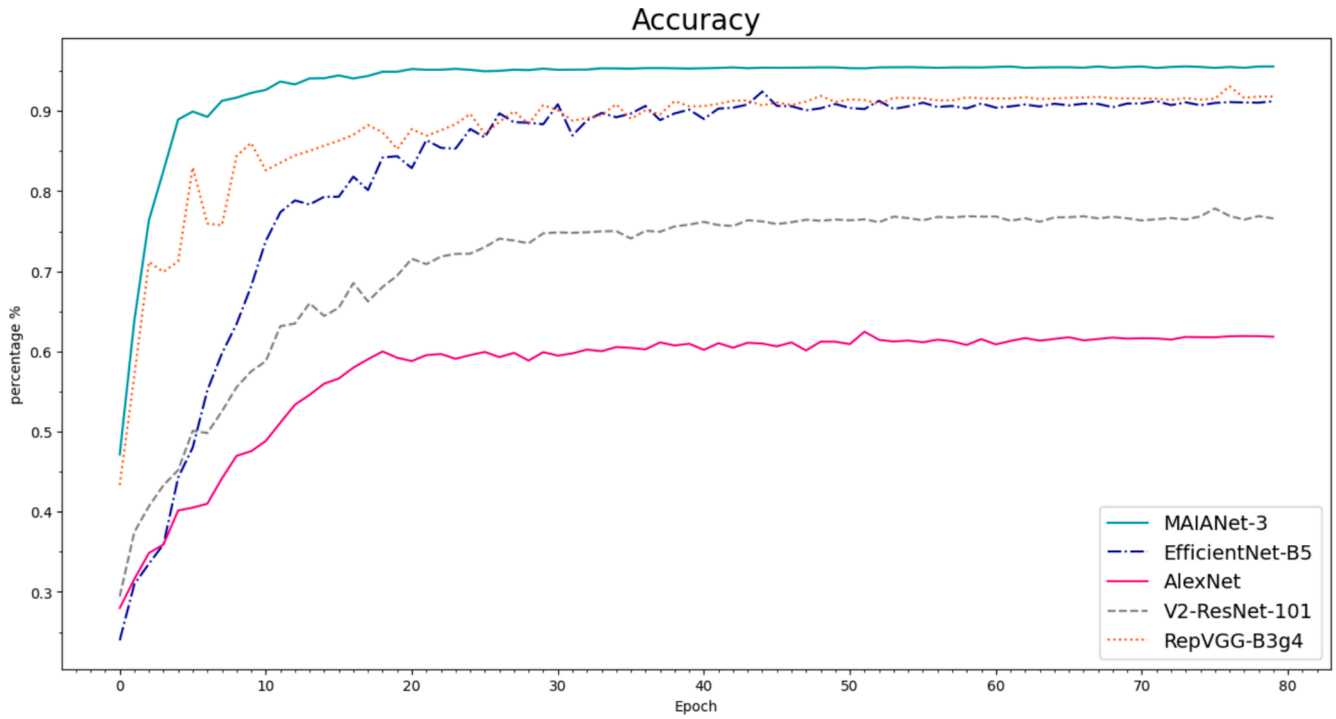


Fig. 6. Accuracy curves of the algorithms on the cassava dataset.

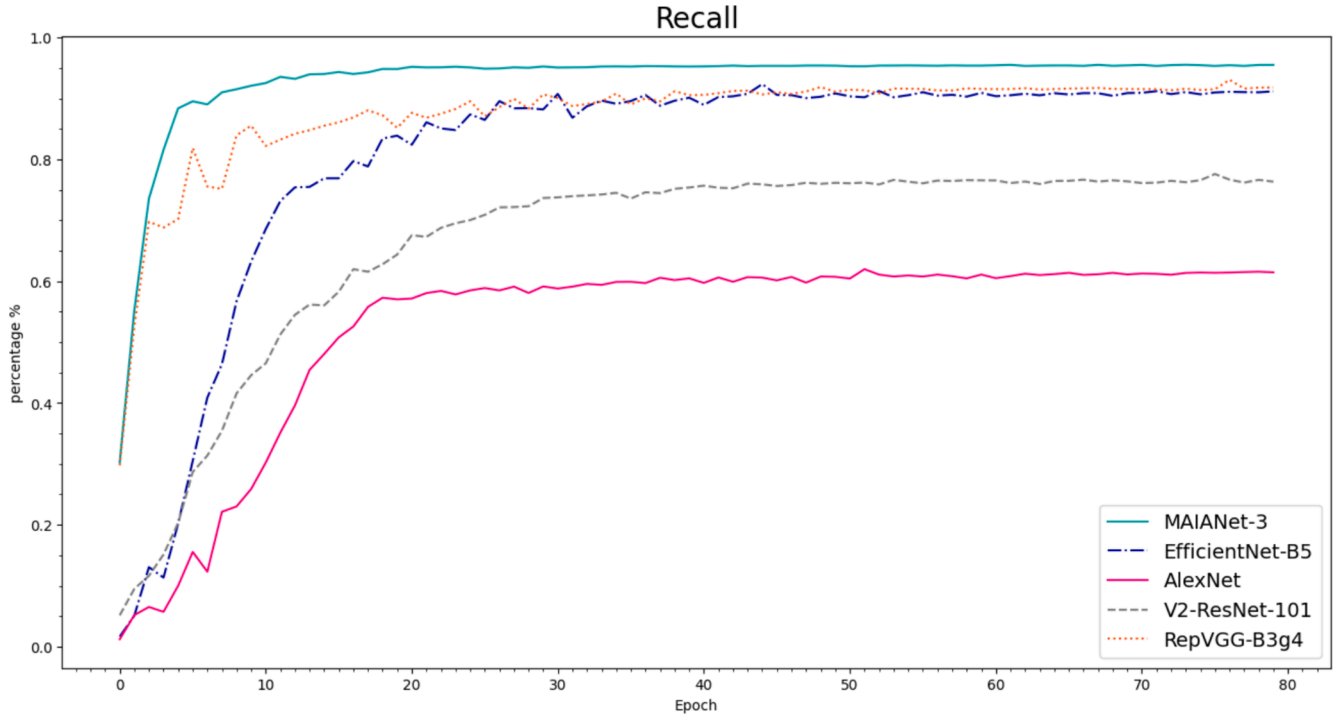


Fig. 7. Recall of the algorithms on the cassava dataset using 7-fold cross validation

unexpected finding is particularly relevant for coarse-grained label-recognition tasks for cassava leaf disease classification. The visualization results indicated that the classification of plant diseases not only relied on outline shapes and infected regions but also required vein information for semantic feature representation. The following visualization results showed that the semantic features of cassava leaves could be fully presented through supplementation of long-distance dependent data.

The partial output feature maps of the residual branch are shown in Fig. 11, and those of the identity branch are shown in Fig. 12. The partial output feature maps of the residual units are shown in Fig. 13. A raw cassava leaf image is shown in Fig. 19 (Appendix C). These visualizations, including Figs. 11, 12, and 13, were observed in the last residual unit of the first stage of the MAIANet-3 neural network.

3.2.1.2. Visualization of the chaotic semantic feature in failed prediction. It was demonstrated that irrelevant features in the input data caused

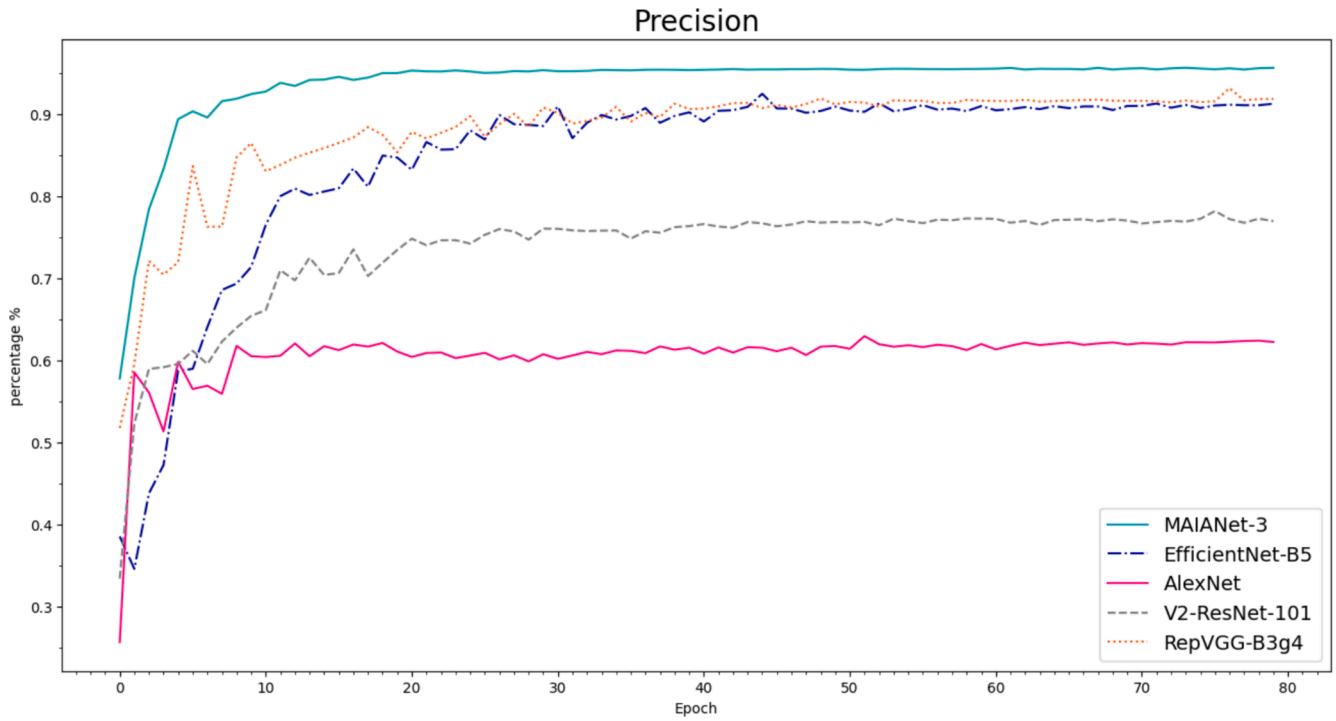


Fig. 8. Precision of the algorithms on the cassava dataset using 7-fold cross validation

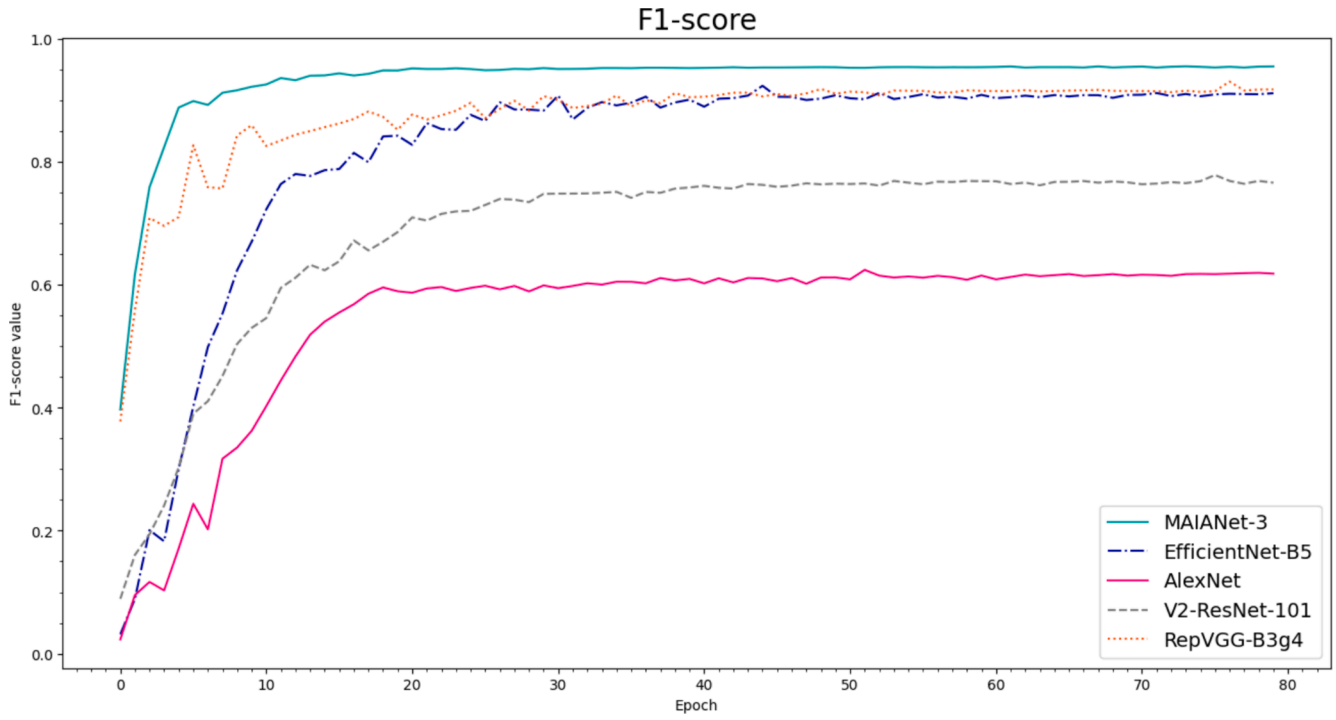


Fig. 9. F1-Score of the algorithms on the cassava dataset using 7-fold cross validation

noise in feature maps. When the descriptor fails to filter out these irrelevant features, it results in disordered semantic features, leading to incorrect predictions. A visualization of the feature maps depicting the failed prediction results is shown in Fig. 14. Irrelevant information is represented as noise in most visualized feature maps. The raw cassava leaf image corresponding to the failed prediction is shown in Fig. 20 (Appendix C). The visualization in Fig. 14 occurs in the last residual unit of the first stage of the MAIANet-3 neural network.

3.3. *Instance batch normalization impacts*

However, as stated by He et al. (2016), the batch normalization layer negatively affects the performance of ResNet when it is used after its residual unit. However, it was found that batch normalization utilized after the residual unit improved the MANet performance. This utilization of batch normalization after the residual unit of MANet formed MABNet.

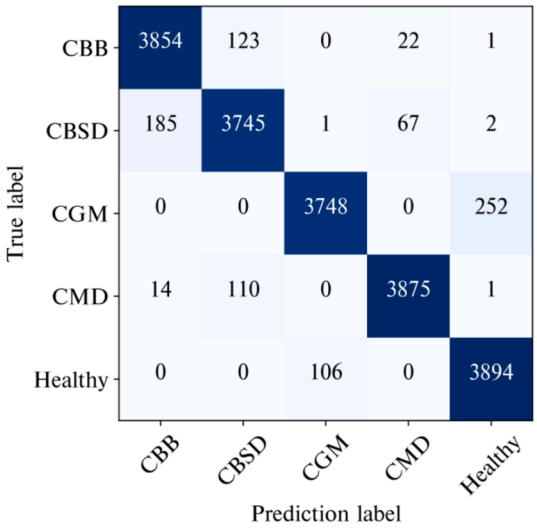


Fig. 10. Confusion matrix of MAIANet-3.

The batch normalization layers significantly improved the performance of the MABNet. Consequently, the instance batch normalization method was introduced to replace the batch normalization layer after the residual unit formed MAIBNet. The performance of the instance batch normalization layer with multiattention is listed in Table 8.

As shown in Table 8, MAIBNet improved the accuracy of MANet by 4.87 % and MABNet by 2.96 %. Furthermore, the F1-score of MAIBNet reached 0.9500, surpassing the F1-score of MABNet by 0.0294 and that of MANet by 0.0487.

As shown in Table 9, the seven-fold cross-validation schedule with a random status of 834,239 demonstrated a superior performance for cassava disease classification. MAIBNet exhibited superior generalization and robustness across different training schemes in the analysis presented in Table 9.

To achieve the highest performance result of the MAIANets, the result of random state 834,239 was used in the following upgrade experiment.

3.4. Anti-aliasing block impacts

An anti-aliasing block was proposed to avoid aliasing signals and represent high-quality semantic features in the sampled feature maps. The number of anti-aliasing blocks was determined based on the feature-signal status of the neural network. Larger feature maps contain more feature information than smaller maps. Thus, the anti-aliasing block was initially utilized in larger feature maps and then used in smaller feature maps until the performance of the neural network decreased. The performance results of MAIANets are shown in Table 10.

As shown by the loss curves in Fig. 15, when the anti-aliasing block continued to be used in stages 1, 2, and 3 of the MAIANet-3 neural network, the loss value improved slightly compared to the MAIANet-2 neural network. The phenomenon of an increase in the loss value was also demonstrated in the MAIANet-4 neural network, and the loss value increased more than in the other comparison algorithms. Furthermore, the performance of MAIANet-4 decreased after the anti-aliasing block was used in all the four stages. This indicates that the feature signal cannot be down-converted in the fourth stage of the MAIANet-3 neural network. When the anti-aliasing block was used in stages 1, 2, and 3 of MAIANet-3, the accuracy of MAIANet-3 reached 95.83 %, an improvement of 0.58 % over MAIBNet.

3.4.1. Visualization of the feature maps of the anti-aliasing block

3.4.1.1. Visualization of the feature maps of the down-conversion block. The input for the depthwise convolution is presented in Fig. 16, whereas the resulting output feature maps of the depthwise convolution are shown in Fig. 17. It is clear that texture and contour information (Nixon and Aguado, 2012) were maintained in the output feature maps. The visualization results of the anti-aliasing block are represented in gray-scale to avoid losing texture information in the visualized result. The visualizations in Figs. 16 and 17 occur in the last residual unit of the first stage of the MAIANet-3 neural network. The feature channel index numbers correspond to those shown in Figs. 16 and 17, respectively.

3.4.1.2. Visualization of the feature maps of the down-sampling operation. Feature maps were visualized using the Matplotlib library (Hunter, 2007). The output feature map of the 2D downsampling convolution operator clearly maintains the semantic information of the coarse-grained labeled input (Fig. 18). The size of the feature map is 56×56.

3.5. Efficiency comparison

To illustrate the computing time of the MAIANet-3 neural network, it was compared to the AlexNet, RepVGG-B3g4, V2-ResNet-101, EfficientNet-B5, MAIBNet, and MANet neural networks. As shown in Table 11, the MAIANet-3 neural network achieved higher accuracy using fewer parameters. Although the computation time increased significantly compared to that of the V2-ResNet-101 neural network, it was shorter than that of the EfficientNet-B5 neural network. The experimental environment presented in Table 11 is detailed in Section 2.3.

To enable a deeper comparison of the computational efficiency of MAIANet-3 and the state-of-the-art (SOTA) neural network, this study

Table 6
Comparison of V2-ResNet-101 with different random status parameters using 7-fold cross validation.

Method	Accuracy	Recall	Precision	F1-score	random status	L2-regularization
V2-ResNet-101	77.84 %	77.61 %	78.17 %	0.7789	834,239	w/o
V2-ResNet-101	86.90 %	86.80 %	87.02 %	0.8692	0	w/o

Annotation: the "w/o" refers to the absence of L2-regularization in the V2-ResNet-101 neural network.

Table 7
MANet vs. other methods on the cassava dataset using 7-fold cross validation.

Method	Accuracy	Recall	Precision	F1-score	Loss	random status	L2-regularization
V2-ResNet-101	86.90 %	86.80 %	87.02 %	0.8692	0.779	0	w/o
SENet (Hu et al., 2018)	83.98 %	83.79 %	84.16 %	0.8398	0.940	0	w/o
FcaNet (Qin et al., 2021)	84.38 %	84.20 %	84.57 %	0.8438	0.939	0	w/o
Double-SENet	89.00 %	88.93 %	89.07 %	0.8900	0.667	0	w/o
MANet	90.12 %	90.00 %	90.28 %	0.9013	0.4653	0	w/o

Annotation: the "w/o" refers to the absence of L2-regularization in the neural network.

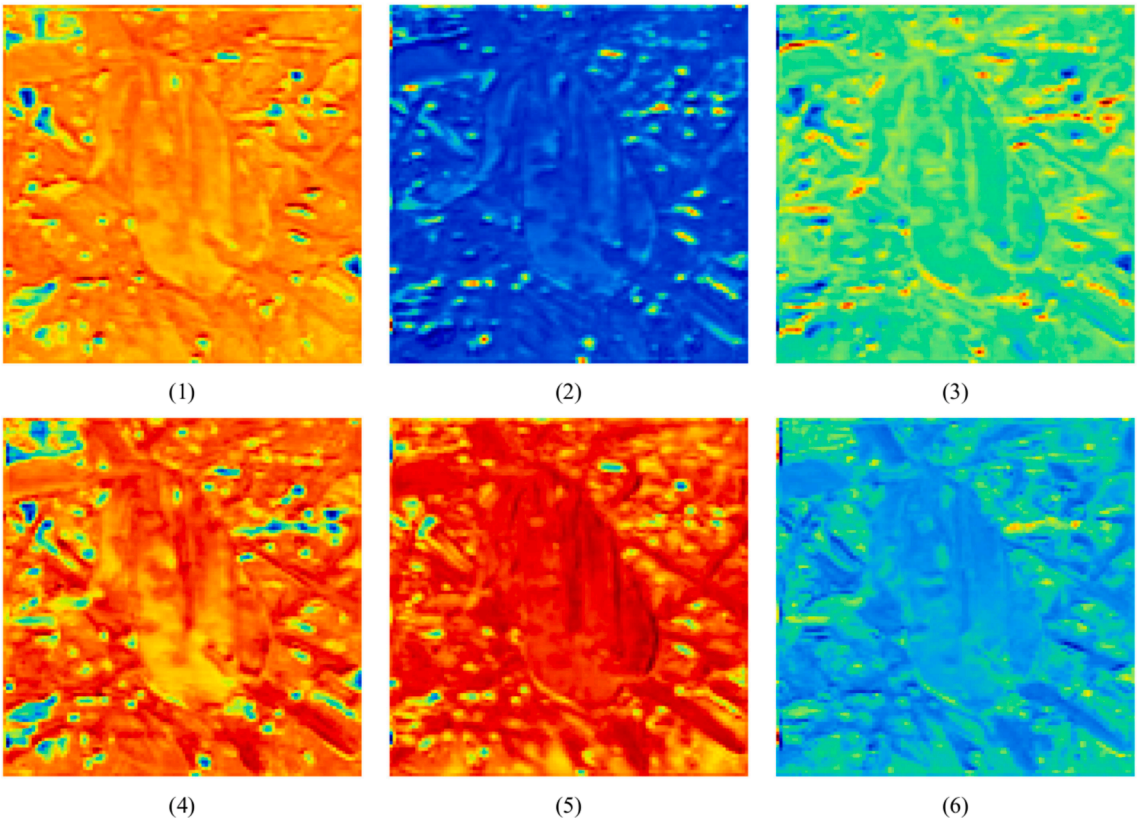


Fig. 11. Part of the output feature maps of the residual branch of the last MAIA block in Stage 1.

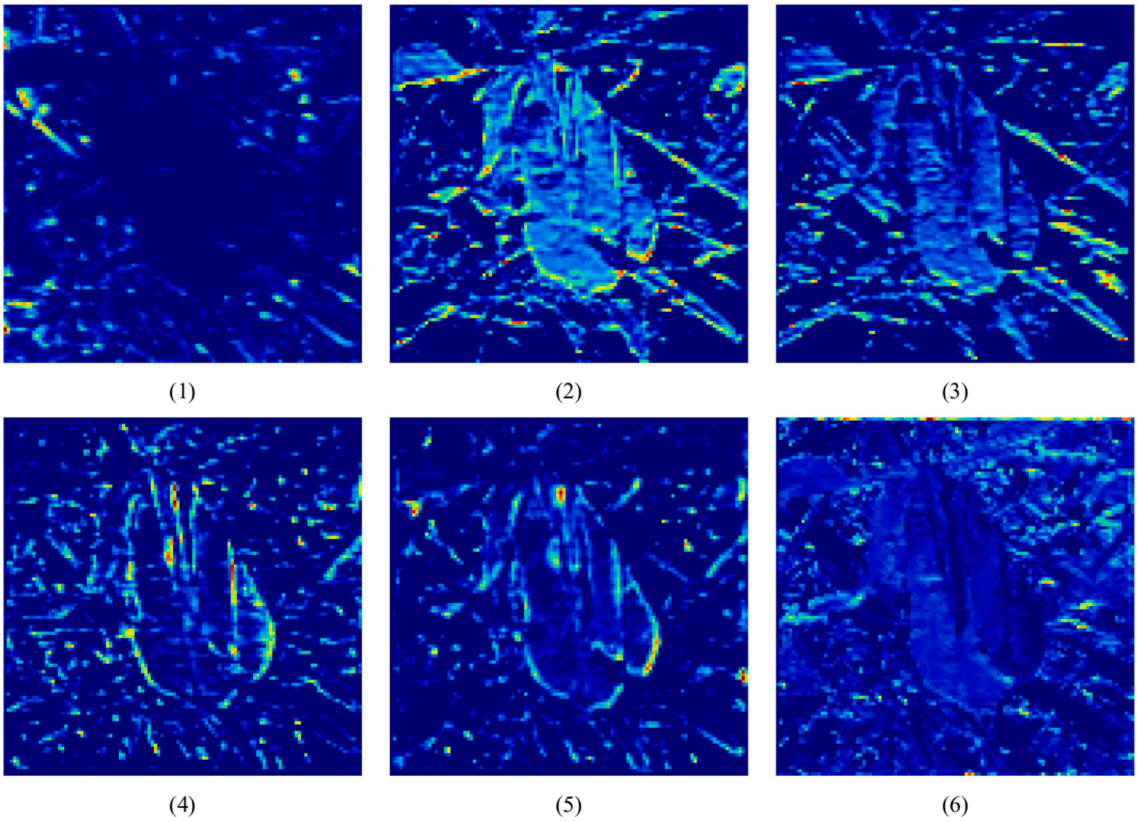


Fig. 12. Part of the feature maps of the identity branch of the last MAIA block in Stage 1.

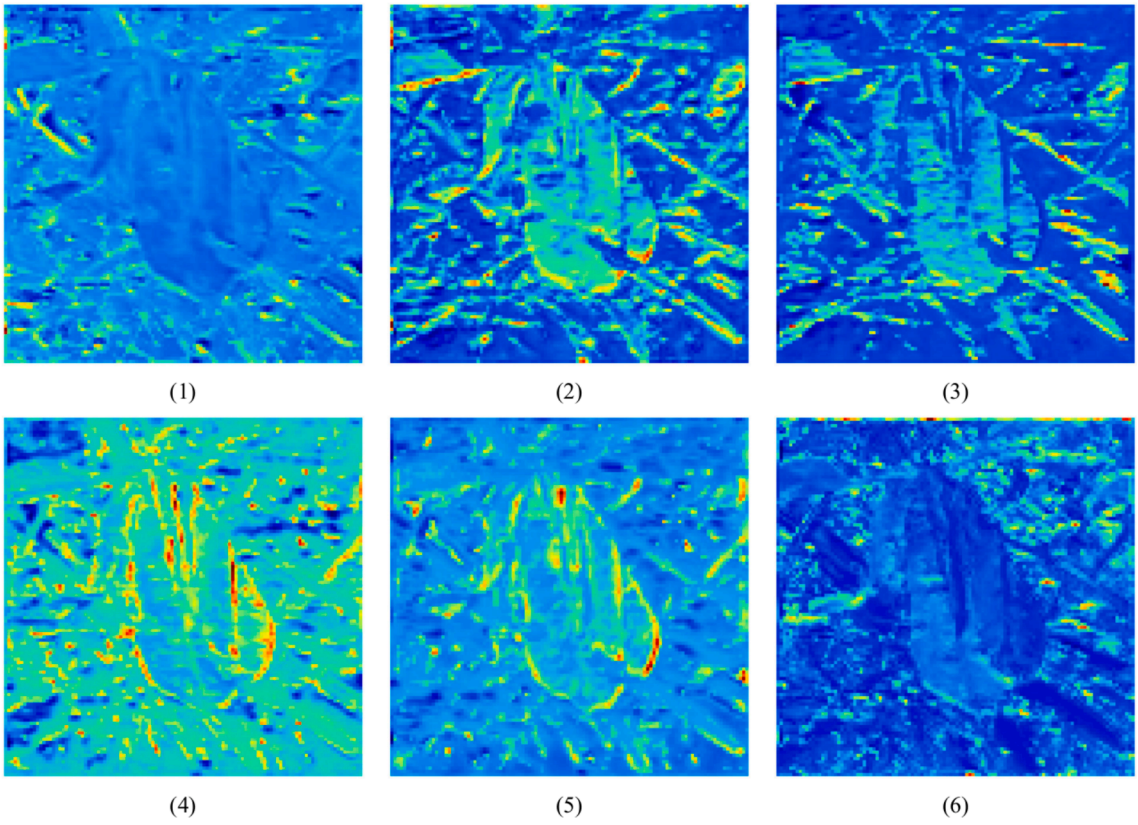


Fig. 13. Part of the output feature maps of the residual unit of the last MAIA block in Stage 1.

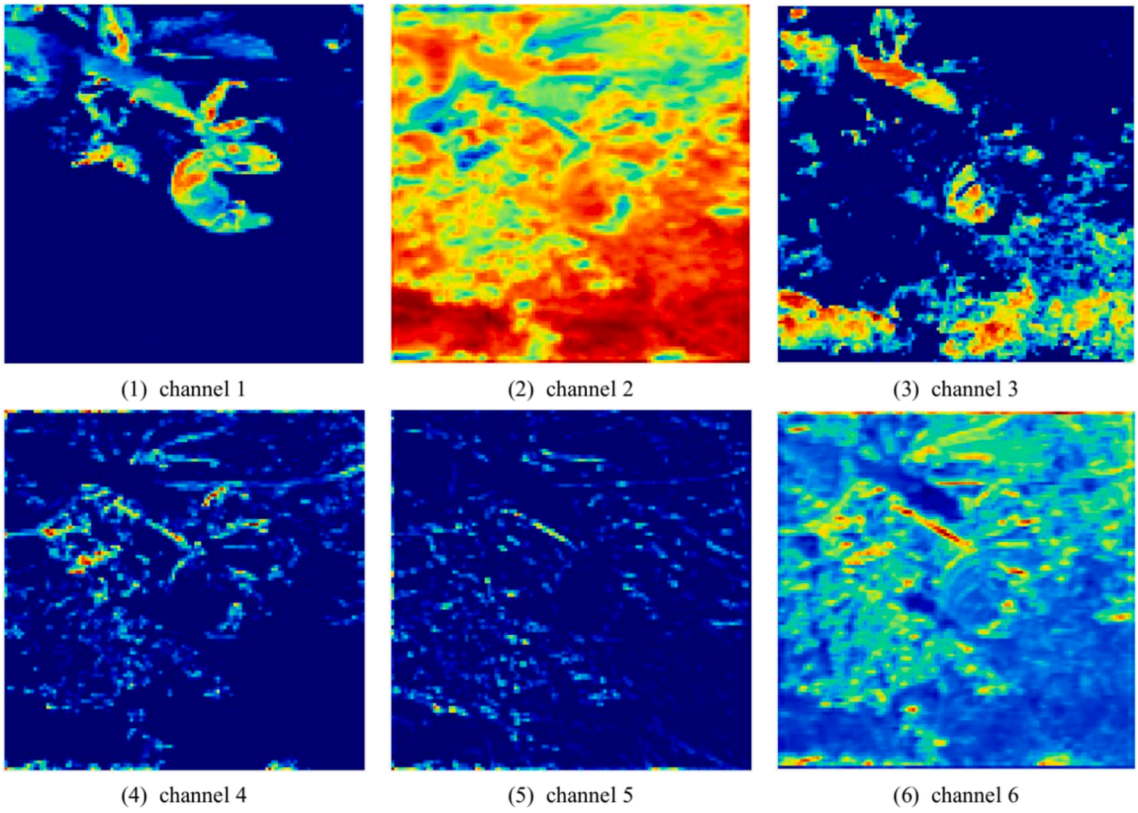


Fig. 14. Part of the visualization feature channels of the prediction failed in the last residual unit of the stage 1.

Table 8
Performance of instance batch normalization layer with multiattention.

Method	Accuracy	Recall	Precision	F1-score	Loss	random status	L2-regularization
MANet	90.12 %	90.00 %	90.28 %	0.9013	0.4653	0	w/o
MABNet	92.03 %	91.98 %	92.13 %	0.9206	0.3783	0	w/o
MAIBNet	94.99 %	94.96 %	95.03 %	0.9500	0.1740	0	w/o

Annotation: the "w/o" refers to the absence of L2-regularization in the MAIANet-3 neural network.

Table 9
Comparison of L2-regularization and random state parameters for MAIBNet and MANet.

Method	Accuracy	Recall	Precision	F1-score	Loss	Random status	L2-regularization
MAIBNet	94.99 %	94.96 %	95.03 %	0.9500	0.1740	0	w/o
	95.17 %	95.11 %	95.20 %	0.9516	0.1840	834,239	w/o
	95.25 %	95.20 %	95.29 %	0.9524	1.6822	834,239	w/

Annotation: the "w/" referred to the utilization of L2-regularization as the optimization method in the MAIANet-3 neural network. In the identity branch, the coefficient was set to 1e-5, and in the residual branch, it was set to 1e-4. The "w/o" indicated the absence of L2-regularization being utilized in the MAIANet-3 neural network.

Table 10
Results of the MAIANet method.

Method	Accuracy	Recall	Precision	F1-score	Loss	anti-aliasing block	random status	L2-RE
MAIBNet	95.25 %	95.20 %	95.29 %	0.9524	1.6822	–	834,239	w/
MAIANet-1	95.26 %	95.22 %	95.30 %	0.9526	3.9454	stage 1	834,239	w/
MAIANet-2	95.74 %	95.72 %	95.80 %	0.9575	1.6901	stage 1, 2	834,239	w/
MAIANet-3	95.83 %	95.81 %	95.82 %	0.9581	1.7190	stage 1, 2, 3	834,239	w/
MAIANet-4	95.72 %	95.70 %	95.77 %	0.9574	1.9179	stage 1, 2, 3, 4	834,239	w/

Annotation: "L2-RE" referred to L2-regularization, and the "w/" indicated the utilization of L2-regularization as the optimization method in the MAIANet-3 neural network. In the identity branch, the coefficient was set to 1e-5, and in the residual branch, it was set to 1e-4.

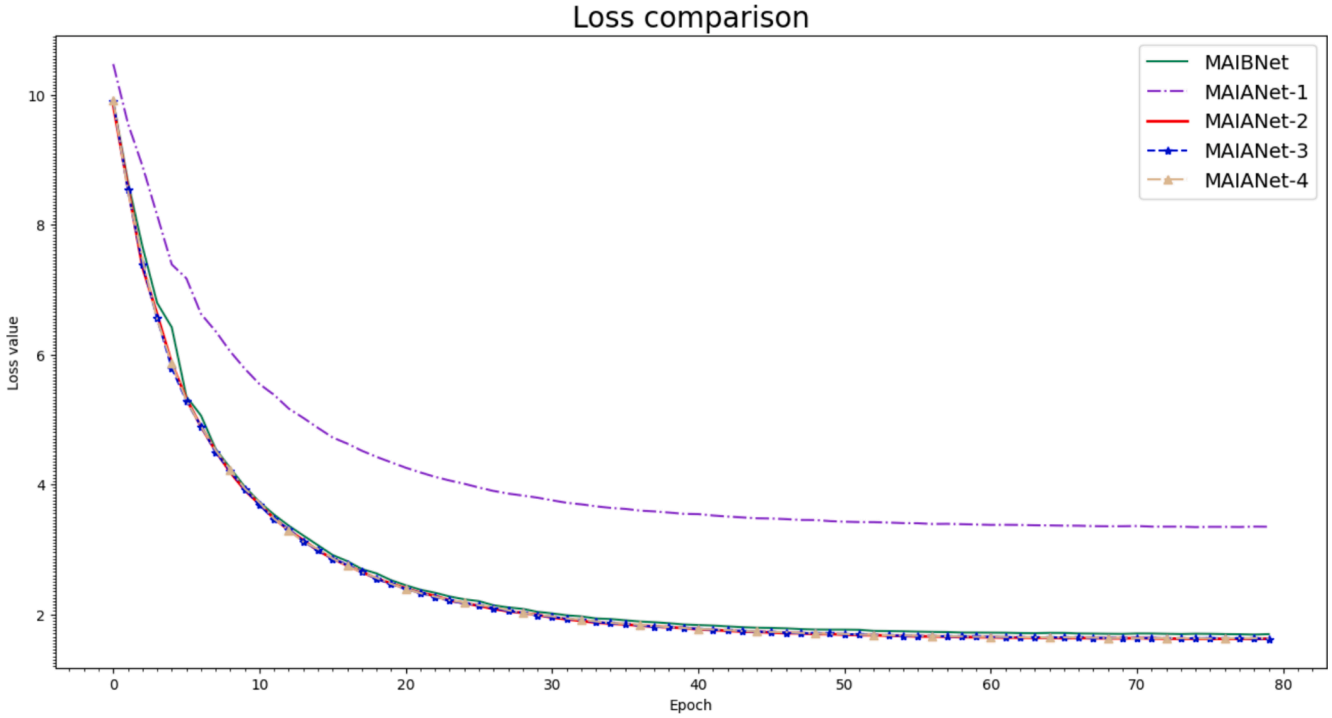


Fig. 15. Comparison of loss curves between MAIANets and MAIBNet.

incorporated the ConvMixer neural network for computational-time benchmarking. The ConvMixer-256/8 (Trockman and Kolter, 2023) was implemented by the Keras team using the Google Colab server with Linux OS. Thus, the training time comparison between MAIANet-3 and ConvMixer was conducted on an Ubuntu 20.04 OS with TensorFlow-GPU 2.8.0 in this study. The training times of the ConvMixer-256/8 and MAIANet-3 neural networks are presented in Table 12.

3.6. Performance of subtle feature identification

This evaluation was conducted based on the manufacturer's data format for the FGVC-Aircraft. To maintain a balanced image distribution, a series of augmentation methods including horizontal flipping, vertical flipping, horizontal-vertical flipping, image offsetting, shift scaling, rotation, and Gaussian noise addition were employed to

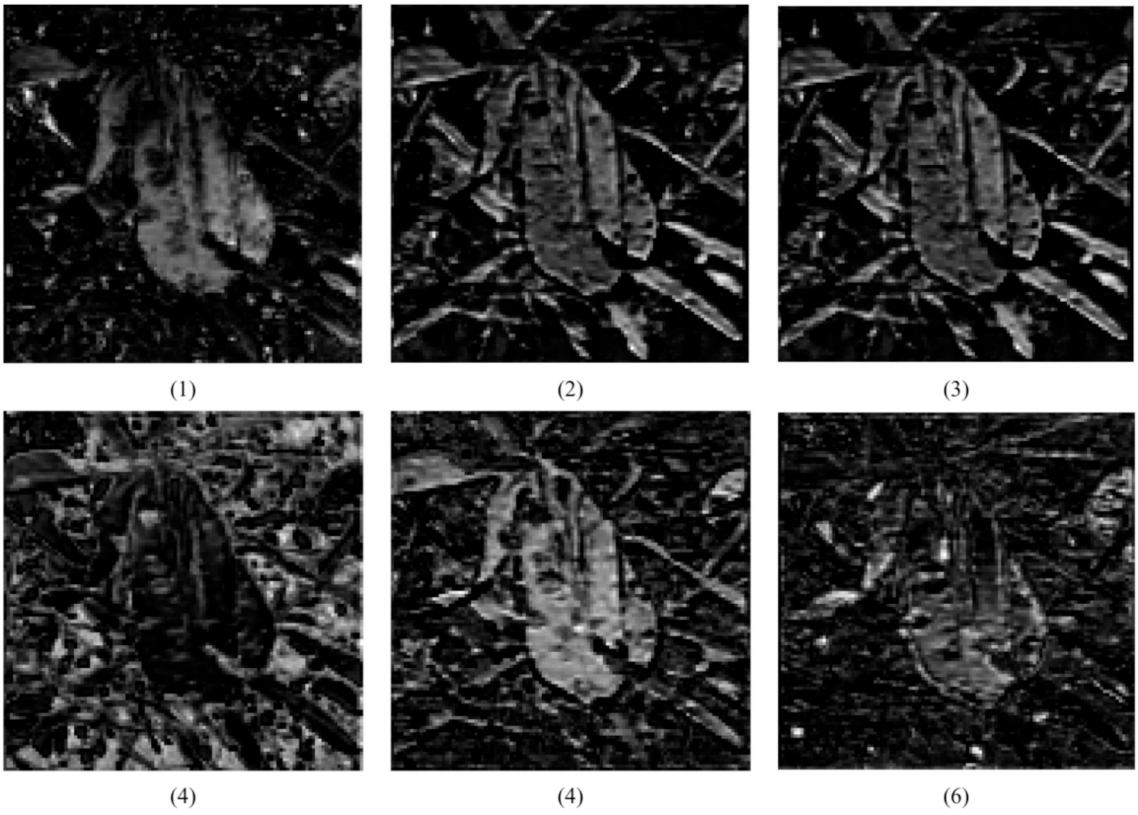


Fig. 16. Part of the input feature maps for the depthwise convolution

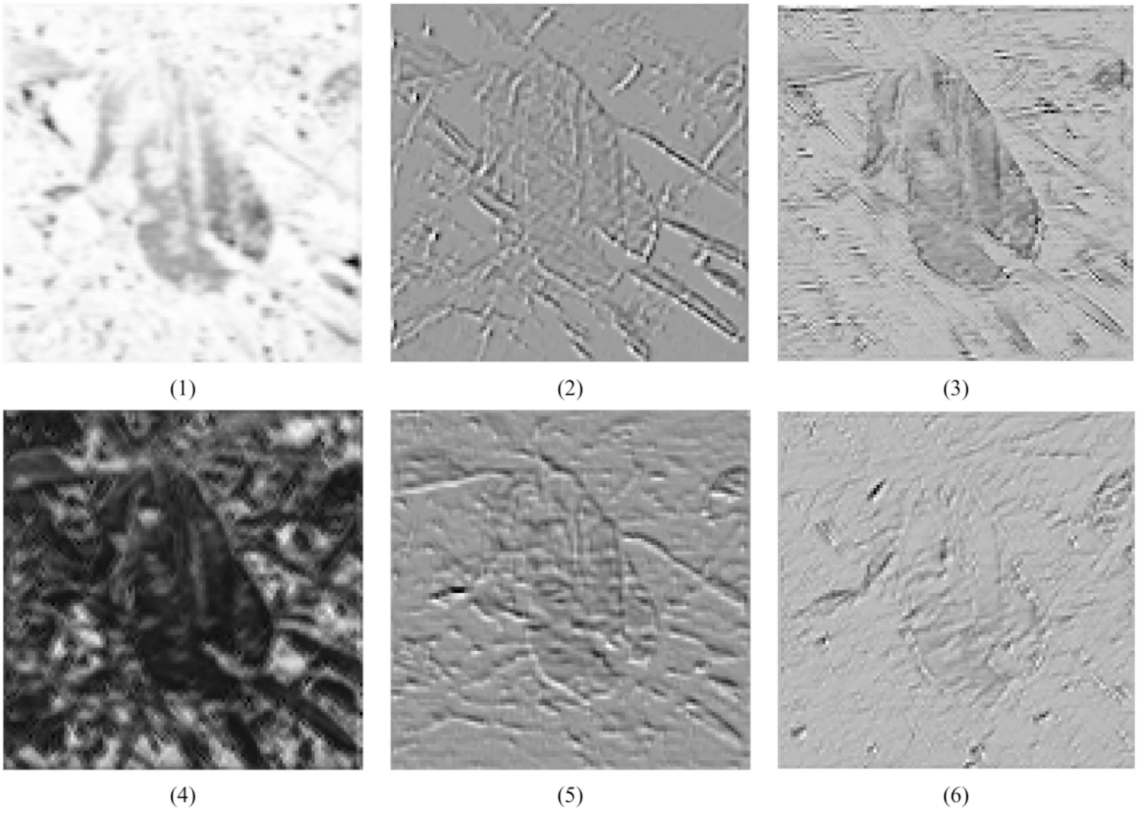


Fig. 17. Part of the output feature maps for the depthwise convolution

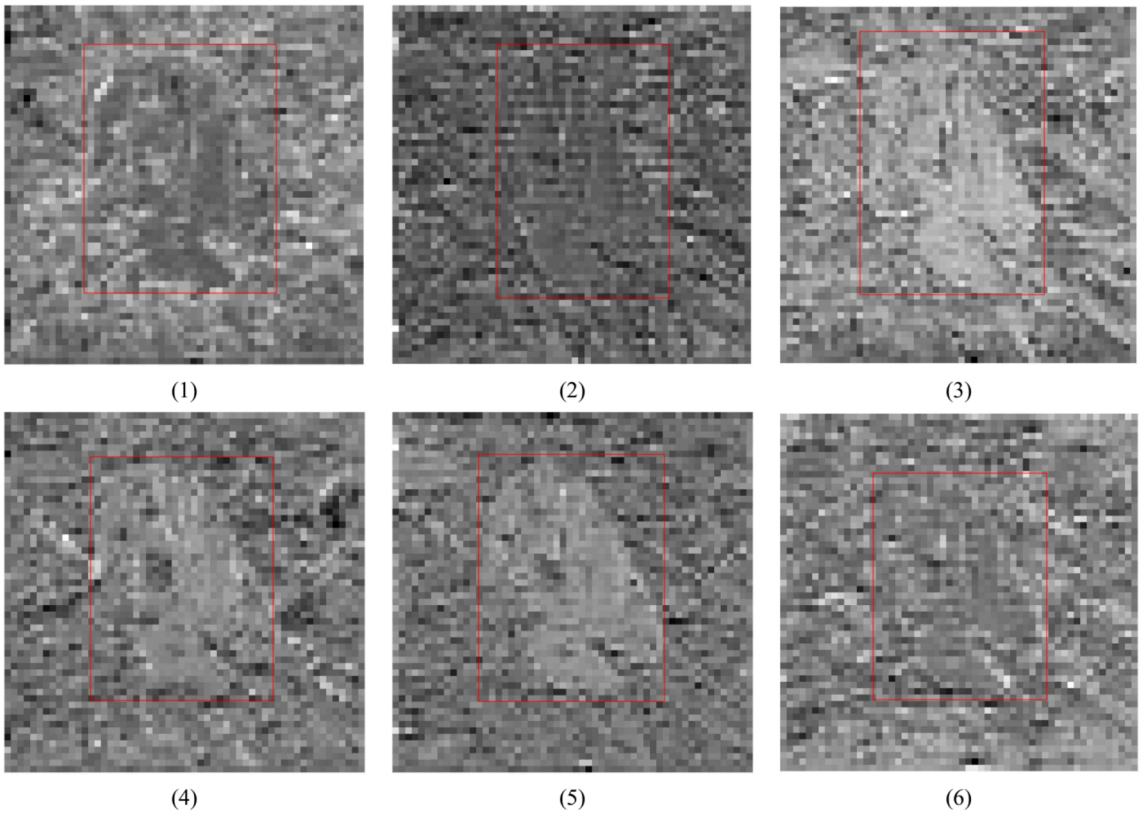


Fig. 18. Part of the output feature maps for the down sampling operation

Table 11
MAIANet-3 vs. other methods based on the comparison of parameters and inference time for Windows 10 OS.

Method (FP 32)	Parameters (M)	Inference time (ms)	Inference speed (images / s)	Train time (minutes / epoch)
MAIANet-3	49.85	18.19	54	16.05
MAIBNet	49.80	17.99	55	15.23
MANet	49.75	14.34	69	9.71
V2-Resnet-101	44.5	6.06	165	6.00
EfficientNet-B5	30	10.01	99	19.32
AlexNet	60	4.93	202	1.68
RepVGG-B3g4	75.62	7.95	125	8.28

increase the number of images. As a result, the dataset comprised 30 categories, with each category containing 1,467 images for training. The training method and parameters are described in detail in Section 2.3.2. The results of the experimental comparisons are presented in Table 13.

MAIANet-3 (based on 50 layers, detailed in Table 14 of Appendix E) achieved an accuracy of 86.19 %, which was 7.39 % higher than the experimental result of Lee et al. (2020) and improved by 3.03 % compared with MAIBNet. The instance batch normalization layer significantly improved the performance of MANet, resulting in an 8.39 % improvement in the accuracy.

Table 12
Training time of the MAIANet-3 on Ubuntu 20.04 OS.

Method (FP 32)	Parameters (M)	Inference time (ms)	Inference speed (images/s)	Train time (minutes / epoch)
MAIANet-3	49.85	12.89	77	12.37
MAIBNet	49.80	12.27	81	11.01
MANet	49.75	9.92	100	8.77
V2-ResNet-101	44.5	5.47	182	5.13
ConvMixer-256/8	0.573735237	25.02	39	45

4. Discussion

Feature maps play a crucial role in the transmission of information within neural networks. To mitigate feature loss during transmission, the multiattention method was utilized to modulate the angle frequency of the feature signals and stack the modulated feature signals into an identity branch to construct the completeness of the semantic feature in the feature maps of the identity branch. As depicted in the visualization of the residual unit within the MAIA block, the vein features were replenished through feature addition operations in the residual units. In addition to the vein features, both the contour and infection features were bolstered by the residual branch. However, vein features can only be effectively extracted in the early stages of a convolutional neural network. As the network deepens, the receptive field expands and the feature map size decreases, leading to the disappearance of the vein

Table 13
Results of our proposed method on FGVC-Aircraft dataset.

Method	Accuracy	Pretrained
ResNet-50 (Lee et al., 2020)	78.8 %	Yes
MAIANet-1 (based 50 layers)	84.25 %	No
MAIANet-2 (based 50 layers)	85.09 %	No
MAIANet-3 (based 50 layers)	86.19 %	No
MAIANet-4 (based 50 layers)	86.10 %	No
MAIBNet (based 50 layers)	83.16 %	No
MANet (based 50 layers)	74.77 %	No

Annotation: the term “based 50 layers” referred to the architecture based on V2-ResNet-50.

feature in the deeper convolution layers.

Multiattention has been proven to modulate critical features to construct high-quality features; however, it is not a permanent channel-attention method. As stated in optical wireless communication systems (OWCS), modulation should be based on one-dimensional feature signals during the frequency modulation of two-dimensional feature signals (Katz and Bar-Ness, 2015; Wang et al., 2002). Therefore, in this study, angle-frequency modulation was established on the one-dimensional feature signal. The enhancement in the classification accuracy of cassava leaf diseases achieved by the multiattention algorithm can be attributed to the precise adjustment of the scale parameter. This adjustment allows the modulated feature signals to exhibit the crucial frequency components necessary for the classification of cassava leaf disease. Furthermore, through ablative experiments conducted on the channel attention algorithm, it became apparent that the variance in performance between the multiattention and double-SE algorithms stemmed from the utilization of both the Fca and SE attention algorithms. Thus, an effective channel attention algorithm should prioritize methods for adjusting the weight parameters based on the input features to generate more accurate weighted coefficients for feature signal modulation. The multiattention method can be reconstructed using different channel-attention methods. With the development of channel attention algorithms, the multiattention approach can be easily modified and improved.

The normalization method effectively corrected the distribution of the eigenvalues in the feature maps; however, it was not a feature descriptor. The instance normalization-corrected features in the maps that batch normalization could not handle in cassava leaf disease classification, aligning with the findings of Wang et al. (2020). Their research suggested that although batch normalization effectively corrected high-frequency component features, it did not significantly enhance network performance when neural networks were trained on input samples containing only low-frequency component features. Thus, it could be inferred that the performance enhancement of instance batch normalization likely stems from its effective correction of low-frequency component feature signals in the feature maps through instance normalization. Additionally, normalization methods have shown notable improvements in generalization performance in subtle feature identification tasks (Jia et al., 2019). Future investigations into the MAIANet-3 neural network aim to incorporate effective normalization methods to further enhance its performance (Choi et al., 2021; Liu et al., 2021a; Zhang et al., 2022b).

The down-conversion block of the anti-aliasing block filtered different frequency component features and re-fitted the high-quality pooling feature maps. However, according to the mathematical expressions of the anti-aliasing block, high-quality pooling feature maps originate from the optimized pooling convolution operator. The optimized pooling convolution operator tended to extract the vein features, contour features, and other texture features, and utilized these features to fit the new feature maps. Therefore, in the pooling operation, vein features, contour features, and other texture features are key features for further improving the semantic feature quality of the pooling feature maps. To further enhance the semantic feature quality in the fourth stage of MAIANet-3, two methods were considered: rotating the numerical values in the numerical space (Zhang et al., 2022a) or modifying the loss function to adjust the feature mapping result.

The presence of soil color information poses a challenge to plant disease classification in complex unstructured environments because it has been mistakenly interpreted as a semantic feature and highlighted in failed prediction feature maps, leading to noise during the cassava leaf semantic extraction process. In the cassava leaf disease images obtained under uncontrollable shooting conditions in the field, due to uneven lighting conditions, soil information may have been clearer than the cassava leaf disease areas, and the soil area may have been larger than the areas affected by cassava leaf lesions. This ultimately led to the neural network model identifying soil color features as cassava leaf disease color features, resulting in errors in neuron parameter fitting and the extraction of incorrect cassava leaf disease features by the neural network model. While color features in the images may not be the most prominent features for cassava leaf disease classification, they were still extracted by the convolutional neural network for fitting cassava leaf disease features and participating in cassava leaf disease categories prediction. This caused errors in neuron parameter fitting, leading the neural network model to extract incorrect cassava leaf disease features. To enhance the accuracy of cassava disease classification, a segmentation method is proposed to discriminate leaf regions and exclude irrelevant areas. Setting a threshold to count the pixels of the lesion area on the leaves could aid in this process, potentially leveraging prior knowledge of the maximum lesion area in certain developmental stages of cassava disease or plant growth, thus facilitating cassava leaf disease prediction based on the weather and cultivar (Gao et al., 2021). Furthermore, augmenting the training datasets with synthetic lesion images can improve the performance of the model. Sun et al. (2020) developed a conventional image processing algorithm to optimize synthetic lesion images obtained from a generative adversarial network (GAN).

The incorporation of the anti-aliasing block resulted in a minor increase in the inference time of MAIANet-3, although it was marginally longer than that of the MAIBNet. It is noteworthy that ConvMixer-256/8, when compared with MAIANet-3, suffered from considerable time complexity, being 3.67 times longer in training time per epoch on a single RTX 3090 GPU. Consequently, ConvMixer-based and transformer neural networks were excluded from the comparative analysis in this study, owing to this significant gap. Among transformer alternatives, the focal modulation network (Yang et al., 2022) demonstrated exceptional performance in classification tasks, suggesting a promising avenue for future neural network design.

The deployment of the proposed neural network model on embedded computation platforms such as NVIDIA Jetson Nano could enable an effective diagnosis of the health status of cassava in the complex unstructured environment of the field. The diagnostic results of cassava could be obtained from a Jetson Nano and integrated into a disease management system. This will enable farmers to implement appropriate treatment plans promptly and provide decision-making information for production management. Moreover, the proposed method can be deployed in agricultural robots or UAV platforms for site-specific disease management. Although satisfactory classification performance was achieved with the open-access dataset, more interesting and challenging work is required to implement the algorithm in a cassava leaf disease system and to accelerate the rapid diagnosis of cassava leaf disease in industrial scenarios. The limitation of our study was the directed classification without excluding irrelevant areas.

Future work will involve further field testing to validate the proposed approach, as well as continued testing of the robustness and accuracy of the proposed model through the addition of new real-world image datasets. The updated model was also used to test images collected at different time points, thereby improving the generalization capabilities of the proposed approach. Additionally, we continuously updated the model using newly labeled datasets and synthetic images to enhance the generalization capabilities. Imaging sensors such as multispectral and hyperspectral cameras have the potential to detect and quantify pre-symptomatic diseases. We continued to optimize the loss functions to improve the computational power of MAIANet-3 while optimizing the normalization methods to further enhance its robustness to domain shifts.

5. Conclusion

This study demonstrated the efficacy of controlling the number of Fourier coefficients when designing a neural network for cassava disease classification. By leveraging theories from Fourier analysis and linear signal systems, we modeled and elucidated the expression of semantic features within CNNs by using the resultant neural network model for cassava disease classification. Through the integration of the

multiattention method, instance batch normalization method, and anti-aliasing block, we introduced a MAIANet-3 neural network tailored for detecting cassava diseases within complex unstructured environments. The manual design of the neural network structure, informed by the derivation of the multiattention and anti-aliasing block formulas, facilitates high-quality frequency-component feature extraction and construction. Based on the visualization of the MAIA block, multiattention has demonstrated impressive performance in constructing veins, contours, and infection features in low-level feature maps. These features are crucial for stacking the complete semantic features within the feature maps of the identity branch, and they are essential for the initial organization of the feature signals in MAIANet-3. Leaf vein, contour, and other texture features were extracted and utilized to reconstruct the completeness of semantic features in pooling feature maps, thereby improving the quality of the feature signals in a convolutional neural network. Furthermore, in this study, we found that leaf vein features were not only the most representative features for classifying plant leaves in simple environments, but also played a role in classifying cassava leaf diseases in complex unstructured environments. However, we simultaneously discovered that leaf contour features, together with leaf vein features, influenced the accuracy of detecting cassava leaf diseases. Additionally, high-frequency component features belonging to both the leaf vein and contour features can be obtained through the modulation of the feature signals in the feature maps.

CRediT authorship contribution statement

Jiayu Zhang: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Baohua Zhang:** Writing – review & editing. **Chao Qi:** Writing – original draft, Data curation. **Innocent Nyalala:** Writing – review & editing. **Peter Mecha:** . **Kunjie Chen:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Junfeng Gao:** Writing – review & editing, Visualization, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Mathematical expression of channel scale

The proof presented in this section relies on the modulation expansion of the pulse signal into the overall signal. The impulse signal within neurons resembles a stimulus signal, triggering the activation of an underlying state. It may not be interpreted strictly as a signal within the brain but rather as the initiation of a memory process within neurons. During the process of neuron activation, the pulse signal amplifies the weak nerve signals in the nerve channel, enabling them to reach the neuron membrane potential, thereby facilitating neuron activation.

Signal modulation frequently relies on complex functions. Integration plays a crucial role in this process, serving as a fundamental mathematical tool with significant implications for signals. Therefore, the modulation-proof process from the pulse signal to the signal was elucidated using integration (Stein and Shakarchi, 2011).

The Poisson summation formula for the objective function of the input information is as follows:

$$\overline{\rho}(t) = \sum_{n=-\infty}^{\infty} \rho(x+nT) = \frac{1}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t},\ \omega_0 = \frac{2\pi}{T} \tag{15}$$

For an impulse signal function, when the maximum value of g(0)→∞, the width of the function infinitely approaches zero.

$$\int_a^b g(t)d_t = K,\ (a\to 0,\ b\to 0,\ and\, a<0<b) \tag{16}$$

Letting $g(t)$ move to the right by $T_0(T_0\to 0)$, obtaining $g(t-T_0)$, the convolution with a shifted impulse formula can be written as follows:

$$\begin{aligned}\overline{\rho}(t)^* g(t-T_0) &= \int_{-\infty}^{\infty} \overline{\rho}(\lambda)g(t-\lambda-T_0)d_\lambda \\ &= \overline{\rho}(t-T_0)\bullet\int_a^b g(t)d_t \\ &= K\bullet\overline{\rho}(t-T_0),\ T_0\to 0\end{aligned} \tag{17}$$

Based on the time delay property of the signal, if $T_0\nrightarrow 0$, the frequency spectrum remains unchanged, and the phase spectrum changes. The equation is as follows:

$$\widehat{\rho}(t-t_0) = P_1(jn\omega_0)e^{jn\omega_0 t} \tag{18}$$

$$P(jn\omega_0) = |P(jn\omega_0)|e^{j\varphi(\omega)},\ P_1(jn\omega_0) = |P_1(jn\omega_0)|e^{jn\omega_0 t} \tag{19}$$

$$\omega t_0 = \varphi(\omega) - \varphi_1(\omega) \tag{20}$$

The offset ωt_0 on the phase spectrum is not required for channel transmission in convolutional neural networks; thus, $T_0\to 0$.

When $K = 1$, $(\rho^* g)(t)$ and is expressed as follows:

$$\overline{\rho}(t)^* g(t-T_0) = \frac{1}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t},\ \omega_0 = \frac{2\pi}{T} \tag{21}$$

When $K = \beta$, $\beta > 1$, $(\rho^* g)(t)$ is expressed as follows:

$$\begin{aligned}\overline{\rho}(t)^* g(t - T_0) &= \frac{\beta}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= \frac{1}{T_\alpha}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega}, \ \omega = \frac{2\pi}{T_\alpha}\end{aligned} \tag{22}$$

where $\frac{\beta}{T} = \frac{1}{T_\alpha}$, $T_\alpha = \frac{T}{\beta}$, because $\frac{2\pi}{\omega_0} = T$, $\frac{1}{T_\alpha} = \frac{\beta\omega_0}{2\pi}$; thus, $\omega = \beta\omega_0$.

When $K = \beta$, $\beta < -1$, $(\rho^* g)(t)$ is expressed as follows:

$$\begin{aligned}\overline{\rho}(t)^* g(t - T_0) &= -\frac{\beta}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= -\frac{1}{T_\alpha}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega t}, \ \omega = \frac{2\pi}{T_\alpha}\end{aligned} \tag{23}$$

When $\beta \to 0$, and $\beta > 0$, assuming $\beta = \frac{1}{\tau}$, and letting $T_\alpha = \tau \mathrm{T}$, the τ is an arbitrarily large value. T_α is the period of the function, and the angular frequency changes to ω. The relationship between the period and the angular frequency is as follows:

$$\omega = \frac{2\pi}{\mathrm{T}_\alpha} \to \omega = \frac{2\pi}{\tau \mathrm{T}} \tag{24}$$

$$\tau\omega = \frac{2\pi}{T} \to \tau\omega = \omega_0 \Longrightarrow \omega = \frac{\omega_0}{\tau} \tag{25}$$

Therefore, $(\rho^* g)(t)$ can be expressed as

$$\begin{aligned}\overline{\rho}(t)^* g(t - T_0) &= \frac{\beta}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= \frac{1}{\tau T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= \frac{1}{T_\alpha}\sum_{n=-\infty}^{\infty} P(jn\omega)e^{jn\omega t}, \ \omega = \frac{2\pi}{T_\alpha}\end{aligned} \tag{26}$$

When $\beta \to 0$, and $\beta < 0$, if $\tau > 0$ and τ refer to an arbitrarily large value, then, $\beta = -\frac{1}{\tau}$, and the proof is as follows:

$$\begin{aligned}\overline{\rho}(t)^* g(t - T_0) &= \frac{\beta}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= \frac{-\frac{1}{\tau} \times 1}{T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= \frac{-1}{\tau T}\sum_{n=-\infty}^{\infty} P(jn\omega_0)e^{jn\omega_0 t} \\ &= -\left\{\frac{1}{T_\alpha}\sum_{n=-\infty}^{\infty} P(jn\omega)e^{jn\omega t}\right\}, \ \omega = \frac{2\pi}{T_\alpha}\end{aligned} \tag{27}$$

In the Cartesian coordinate system, flipping around the X-axes can be considered when a negative sign on the one-dimensional function is outside the system. The high-frequency signals in the spectrum of the feature signal converge towards the origin, and the convolution function can obtain more high-frequency signals. Consequently, high-frequency signals are maintained in the channels of the convolutional neural network.

Appendix B. Mathematical expression of the semantic feature deconstruction

Depthwise convolution was used to down-convert the frequency of the feature map. The proof is as follows:

When the convolution kernel is in its initial state,

$$\delta = f(x) - \int_{-\infty}^{\infty} f(x)g(y - x)d_x \tag{28}$$

where $f(x)$ refers to the objective information of the input signals, and $g(y - x)$ refers to the initial kernel. δ refers to the information loss. When the kernel g approaches a good kernel state, the expression can be written as

$$v = f(x) - \int_{-\infty}^{\infty} f(x)g(y - x)d_x \tag{29}$$

As the integrity of the input signal increased, the kernel function g referred to the destination state. v refers to information loss, the value of which

gradually decreases as the number of training epochs increases. The backpropagation mechanism attempts to increase the value of τ as follows:

$$\delta - \nu = \tau \tag{30}$$

In the ideal state, when the kernel function g refers to a good kernel $\nu = 0$, and$\tau = \delta$. However, it is difficult to implement $\nu = 0$ in local optimum gradient. Thus, the $\omega_{kernel} \neq \omega_{signal}$, can be expressed as follows:

$$\alpha \bullet \omega_{kernel} = \beta \bullet \omega_{signal} \ \left(\omega_{signal} \to 0,\ \alpha \in C,\ and\, \beta \to \infty\right) \tag{31}$$

$$\omega_p = \frac{\alpha}{\beta}\omega_{kernel} = \omega_{signal},\ \frac{\alpha}{\beta} \to 0 \tag{32}$$

where α and β refer to $\mathbb{R}^+$; ω_{signal} is the angular frequency of the input signal; ω_{kernel} is the angular frequency of the convolution kernel; and ω_p is the angular frequency of the convolution result. The value of $\frac{\beta}{\alpha}$ was adjusted using the back-propagation method. The feature map signal is an aperiodic continuous spectrum signal; therefore, $\omega_{signal} \to 0$.

where $\frac{\alpha}{\beta} \to 0$ is the ideal mathematical expression. ω_{kernel} prefer to maintain the same angle-frequency component signals in the convoluted results. Loss of signal frequency was inevitable, and it was impossible to determine the lost signal frequency component, which was determined by the gradient orientation.

Appendix C. Raw cassava leaf disease image

Fig. 19. Disease cassava leaf image with background noise.

Fig. 20. Original image of cassava leaf disease that failed to be detected.

Appendix D. Fourier coefficient

The Fourier coefficient can be calculated using Fourier transform. The Fourier coefficient of the periodic discrete-time signal is given by Equation (33):

Any periodic discrete-time signal with a period N can be expressed as a linear combination of N exponential functions (Lai, 2003). Given a signal $x(n)$, the Fourier coefficient can be calculated using the following equation:

$$C_k = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-j2\pi kn/N} \tag{33}$$

The Fourier coefficient is an essential property of Fourier analysis. In the frequency domain, a real number is split and distributed to numerous angular frequencies formed by complex numerical values as a Fourier coefficient, which is the concept of Fourier transform. The inverse Fourier transform was converted and summed from the Fourier coefficients of the various angular frequencies to a real number. In communication theory, the number of Fourier coefficients corresponds to signal quality and integrity. Thus, the frequency component is essential for the signal representation quality.

Appendix E. Maianet architecture

Table 14
Architecture of MAIANet-3 for cassava leaf disease classification.

layer name	output size	50-layer	101-layer
conv1	224 × 224	7 × 7, 64, stride 2	
	112 × 112	3 × 3 max pool, stride 2	
conv2_x	112 × 112	$\text{Anti}-\text{aliasing block}: \left\{ \begin{bmatrix} 1\times1 & str:1 & 64 \\ Down-conversionblock & & \\ 3\times3 & str:1 & 64 \\ 1\times1 & str:1 & 256 \end{bmatrix}_{Residual} \right\} \times 1$	
		$\left\{ \begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 2$	$\left\{ \begin{bmatrix} 1\times1, & 64 \\ 3\times3, & 64 \\ 1\times1, & 256 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 2$
conv3_x	56 × 56	$\text{Anti}-\text{aliasing block}: \left\{ \begin{bmatrix} 1\times1 & str:1 & 128 \\ Down-conversionblock & & \\ 3\times3 & str:2 & 128 \\ 1\times1 & str:1 & 512 \end{bmatrix}_{Residual} \right\} \times 1$	
		$\left\{ \begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 3$	$\left\{ \begin{bmatrix} 1\times1, & 128 \\ 3\times3, & 128 \\ 1\times1, & 512 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 3$
conv4_x	28 × 28	$\text{Anti}-\text{aliasing block}: \left\{ \begin{bmatrix} 1\times1 & str:1 & 256 \\ Down-conversionblock & & \\ 3\times3 & str:2 & 256 \\ 1\times1 & str:1 & 1024 \end{bmatrix}_{Residual} \right\} \times 1$	
		$\left\{ \begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 5$	$\left\{ \begin{bmatrix} 1\times1, & 256 \\ 3\times3, & 256 \\ 1\times1, & 1024 \\ MAblock & \end{bmatrix}_{Resdiual} \right\} \times 22$

(continued on next page)

Table 14 (*continued*)

layer name	output size	50-layer	101-layer
conv5_x	14 × 14	$\left\{\begin{bmatrix}1\times1 & str:1 & 512\\ 3\times3 & str:2 & 512\\ 1\times1 & str:1 & 2048\end{bmatrix}_{Residual}\right\}\times1$	
		$IBN(add(Residual, Identity))$	
		$\left\{\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\\ MAblock & \end{bmatrix}_{Resdiual}\right\}\times2$	$\left\{\begin{bmatrix}1\times1, & 512\\ 3\times3, & 512\\ 1\times1, & 2048\\ MAblock & \end{bmatrix}_{Resdiual}\right\}\times2$
		$IBN(add(Residual, Identity))$	$IBN(add(Residual, Identity))$
Prediction	1 × 1	Global average pooling 2D, 5-d fc, Softmax activation	

Appendix F. Projection distance relation to angle frequency and Fourier coefficient

A projection distance exists between the fitting result and the true result. In the Fourier analysis, this method is used to pursue the optimal approximation theorem for the projection distance. The projection distance was calculated using Equation (34).

$$\|f\|^2 = \|f - S_n(f)\|^2 + \|\sum_{|n|\le N}|a_n|^2,\ S_N(f) = \sum_{|n|\le N}a_n e_n \tag{34}$$

where f is an integrable function on the circle, a_n is the Fourier coefficient of f, n refers to the integer numerical value, $\|\partial\|$ refers to the norm of ∂, and c_n refers to the assumed Fourier coefficient, and $c_n = a_n - b_n$, $e_n(\theta) = e^{in\theta}$, and the list $\{e_n\}_{n\in\mathbb{Z}}$ refers to the standard norm element list. The final mathematical expression for Equation (34) is given by Equation (35).

$$Thus:\ f - \sum_{|n|\le N}c_n e_n = f - S_N(f) + \sum_{|n|\le N}b_n e_n \tag{35}$$

Let norm $\|f\|^2 \to \|f\|$, Equation (35) is derived to Equation (36). Equation (36) is the optimal approximation theorem of the Fourier analysis.

$$\|f - S_N(f)\| \le \|f - \sum_{|n|\le N}c_n e_n\| \tag{36}$$

where, Equation (36) is satisfied with any complex numerical value of c_n. When $c_n = a_n$, both sides of Equation (36) are equal. The minimum value of $\|f - S_N(f)\|$ is found in the frequency of the N−th trigonometric polynomial. The equation $b_n = a_n - c_n$ is the essential element in Equation (36), which is related to the angle frequency and value of the Fourier coefficient a_n. The geometric interpretation of the best approximation lemma of Fourier analysis is shown in Fig. 21.

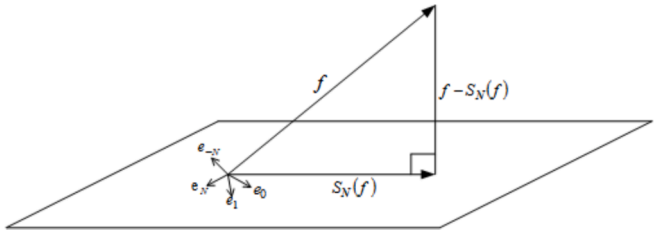


Fig. 21. Geometric interpretation of the best approximation lemma of Fourier analysis.

Appendix G. Performance exhibition of MAIANet-3 without L2-regularization

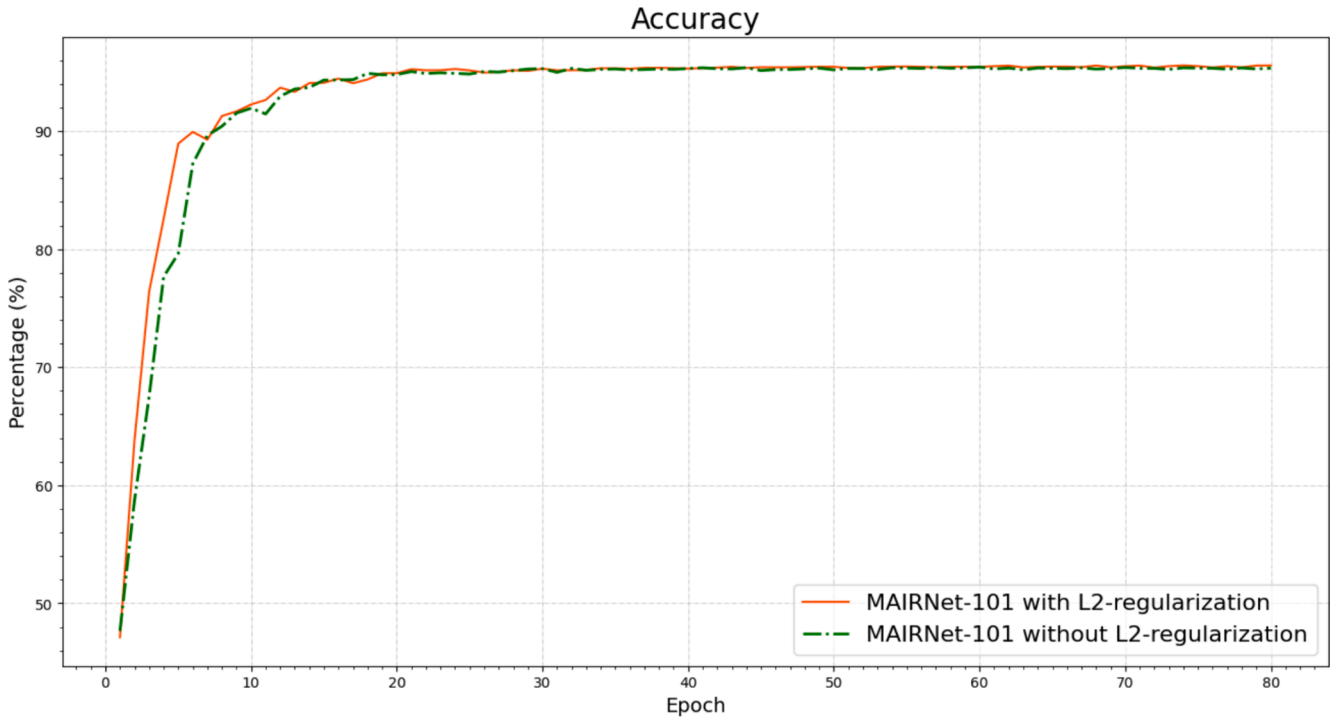


Fig. 22. Comparison of MAIANet-3 accuracy curves with and without L2-regularization.

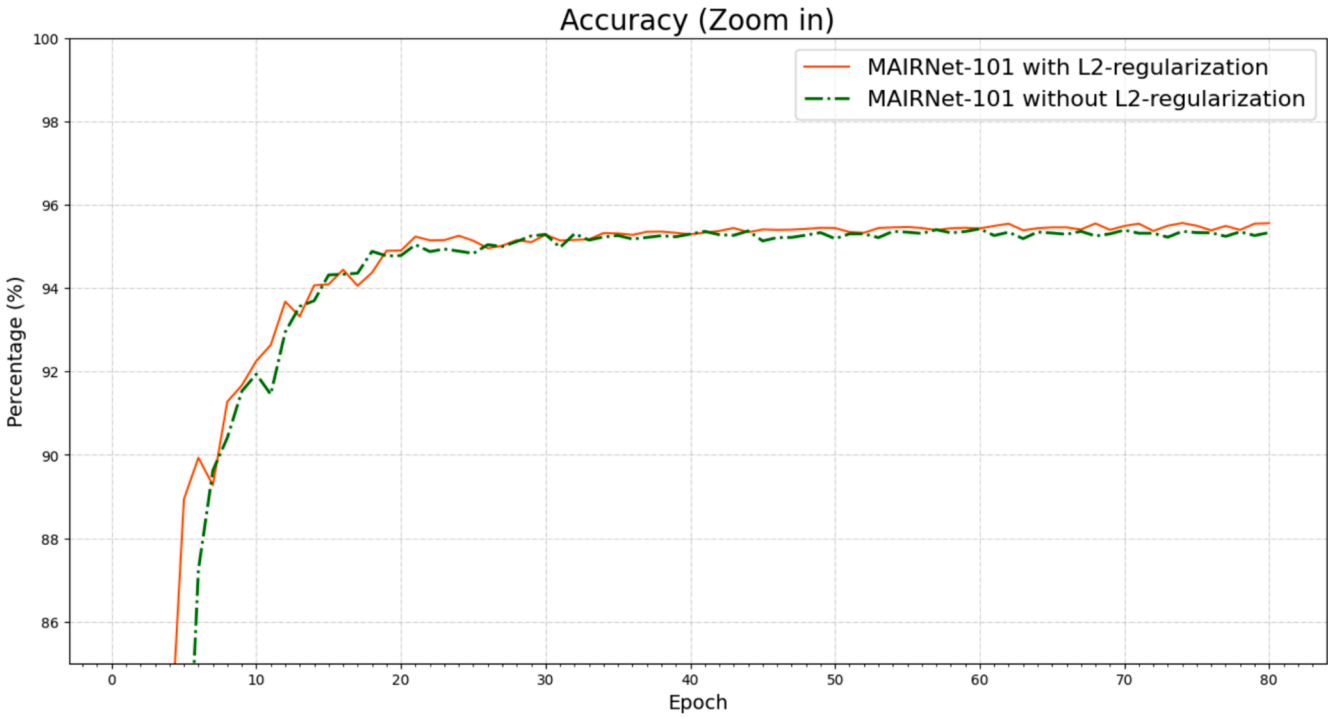


Fig. 23. Detailed Comparison of MAIANet-3 accuracy curves with and without L2-regularization

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: a system for Large-Scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.

Abdullahi, I., Atiri, G., Dixon, G., et al., 2003. Effects of cassava genotype, climate and the Bemisia tabaci vector population on the development of African cassava mosaic geminivirus (ACMV). Acta Agron. Hung. 51, 37–46.

Azeroual, O., 2020. Data wrangling in database systems: purging of dirty data. Data 5, 50.
BS, P., 2022. Disease Classification and Detection Techniques in Rice Plant using Deep Learning, 2022 8th International Conference on Smart Structures and Systems (ICSSS), pp. 1-7.
Chang, D., Ding, Y., Xie, J., Bhunia, A.K., Li, X., Ma, Z., Wu, M., Guo, J., Song, Y.-Z., 2020. The devil is in the channels: Mutual-channel loss for fine-grained image classification. IEEE Trans. Image Process. 29, 4683–4695.
Chen, Z., Chen, J., Feng, Y., Liu, S., Zhang, T., Zhang, K., Xiao, W., 2022. Imbalance fault diagnosis under long-tailed distribution: Challenges, solutions and prospects. Knowl.-Based Syst. 258, 110008.
Chen, J., Deng, X., Wen, Y., Chen, W., Zeb, A., Zhang, D., 2023. Weakly-supervised learning method for the recognition of potato leaf diseases. Artif. Intell. Rev. 56, 7985–8002.
Chisenga, S.M., Workneh, T.S., Bultosa, G., Alimi, B.A., 2019. Progress in research and applications of cassava flour and starch: a review. J. Food Sci. Technol. 56, 2799–2813.
Choi, S., Kim, T., Jeong, M., Park, H., Kim, C., 2021. Meta batch-instance normalization for generalizable person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3425–3435.
Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258.
Dhivyaa, C.R., Kandasamy, N., Rajendran, S., 2022. Integration of dilated convolution with residual dense block network and multi-level feature detection network for cassava plant leaf disease identification. Concurrency and Computation: Practice and Experience 34, e6879.
Ding, X., Zhang, X., Ma, N., Han, J., Ding, G., Sun, J., 2021. Repvgg: Making vgg-style convnets great again. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13733–13742.
Elango, F., Lozano, J., 1980. Transmission of Xanthomonas manihotis in seed of cassava (Manihot esculenta). Plant Dis 64, 784–786.
Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318.
Gao, S.-H., Cheng, M.-M., Zhao, K., Zhang, X.-Y., Yang, M.-H., Torr, P., 2019. Res2net: A new multi-scale backbone architecture. IEEE Trans. Pattern Anal. Mach. Intell. 43, 652–662.
Gao, J., Westergaard, J.C., Sundmark, E.H.R., Bagge, M., Liljeroth, E., Alexandersson, E., 2021. Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. Knowl.-Based Syst. 214, 106723.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks, Computer Vision–ECCV 2016. In: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer, pp. 630–645.
Hillocks, R., Raya, M., Thresh, J., 1996. The association between root necrosis and above-ground symptoms of brown streak virus infection of cassava in southern Tanzania. International Journal of Pest Management 42, 285–289.
Hillocks, R.J., Thresh, J., 2002. Cassava: biology, production and utilization. CABI Publishing.
Howeler, R., Lutaladio, N., Thomas, G., 2013. Save and grow: cassava: a guide to sustainable production intensification. Food and Agriculture Organization of the United Nations Rome.
Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95.
Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning. Pmlr 448–456.
Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 20, 1254–1259.
Jerripothula, K.R., Cai, J., Yuan, J., 2016. Cats: Co-saliency activated tracklet selection for video co-localization, Computer Vision–ECCV 2016. In: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VII 14. Springer, pp. 187–202.
Jia, J., Ruan, Q., Hospedales, T.M., 2019. Frustratingly easy person re-identification: Generalizing person re-id in practice. arXiv preprint arXiv:1905.03422.
Jin, C., Schenkel, M., Carlile, S., 2000. Neural system identification model of human sound localization. J. Acoust. Soc. Am. 108, 1215–1235.
Katz, E., Bar-Ness, Y., 2015. Two-dimensional (2-D) spatial domain modulation methods for unipolar pixelated optical wireless communication systems. J. Lightwave Technol. 33, 4233–4239.
Koch, C., Ullman, S., 1987. Shifts in selective visual attention: towards the underlying neural circuitry, Matters of intelligence: Conceptual structures in cognitive neuroscience. Springer, pp. 115–141.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Proces. Syst. 25.
Lai, E., 2003. Practical digital signal processing. Elsevier.
Lee, S.H., Chan, C.S., Mayo, S.J., Remagnino, P., 2017. How deep learning extracts and learns leaf features for plant classification. Pattern Recogn. 71, 1–13.
Lee, J., Won, T., Lee, T.K., Lee, H., Gu, G., Hong, K., 2020. Compounding the performance improvements of assembled techniques in a convolutional neural network. arXiv preprint arXiv:2001.06268.
Legg, J.P., Kumar, P.L., Makeshkumar, T., Tripathi, L., Ferguson, M., Kanju, E., Ntawuruhunga, P., Cuellar, W., 2015. Cassava virus diseases: biology, epidemiology, and management. Advances in Virus Research. Elsevier 85–142.
Li, Y.-F., Guo, L.-Z., Zhou, Z.-H., 2019. Towards safe weakly supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. 43, 334–346.
Lilhore, U.K., Imoize, A.L., Lee, C.-C., Simaiya, S., Pani, S.K., Goyal, N., Kumar, A., Li, C.-T., 2022. Enhanced convolutional neural network model for cassava leaf disease identification and classification. Mathematics 10, 580.
Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988.
Liu, X., Min, W., Mei, S., Wang, L., Jiang, S., 2021. Plant disease recognition: A large-scale benchmark dataset and a visual region and loss reweighting approach. IEEE Trans. Image Process. 30, 2003–2015.
Liu, S., Zhang, K.-Y., Yao, T., Bi, M., Ding, S., Li, J., Huang, F., Ma, L., 2021a. Adaptive normalized representation learning for generalizable face anti-spoofing. In: Proceedings of the 29th ACM International Conference on Multimedia, pp. 1469–1477.
Loshchilov, I., Hutter, F., 2019. Decoupled weight decay regularization. International Conference on Learning Representations.
Ma, J., Du, K., Zheng, F., Zhang, L., Gong, Z., Sun, Z., 2018. A recognition method for cucumber diseases using leaf symptom images based on deep convolutional neural network. Comput. Electron. Agric. 154, 18–24.
Maji, S., Rahtu, E., Kannala, J., Blaschko, M., Vedaldi, A., 2013. Fine-grained visual classification of aircraft. arXiv preprint arXiv:1306.5151.
Makerere Artificial Intelligence Laboratory. (February 19, 2021). Cassava leaf disease dataset. https://www.kaggle.com/competitions/cassava-leaf-disease-classification.
Maraite, H., 1993. Xanthomonas campestris pathovars on cassava: cause of bacterial blight and bacterial necrosis. Xanthomonas 18–25.
Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., 2015. Advanced methods of plant disease detection. A Review. Agronomy for Sustainable Development 35, 1–25.
McCallum, E.J., Anjanappa, R.B., Gruissem, W., 2017. Tackling agriculturally relevant diseases in the staple crop cassava (Manihot esculenta). Curr. Opin. Plant Biol. 38, 50–58.
Moratal, D., Vallés-Luch, A., Martí-Bonmatí, L., Brummer, M.E., 2008. k-Space tutorial: an MRI educational tool for a better understanding of k-space. Biomed. Imaging Intervention J. 4.
Ng, A.Y., 2004. Feature selection, L 1 vs. L 2 regularization, and rotational invariance, Proceedings of the twenty-first international conference on Machine learning, p. 78.
Nixon, M.S., Aguado, A.S., 2012. Feature extraction & image processing for computer vision. Academic press.
NVIDIA. (2023). ResNet v1.5 for PyTorch. NVIDIA NGC: AI Development Catalog. Retreved 2023/9/3 from https://ngc.nvidia.com/catalog/model-scripts/nvidia:resnet_50_v1_5_for_pytorch.
Oyewola, D.O., Dada, E.G., 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.
Pan, X., Luo, P., Shi, J., Tang, X., 2018. Two at once: Enhancing learning and generalization capacities via ibn-net. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 464–479.
Paper with code.(Apr. 27, 2024). Statistical of FGVC-Aircraft dataset. https://paperswithcode.com/dataset/fgvc-aircraft-1.
Polder, G., van der Heijden, G.W., van Doorn, J., Baltissen, T.A., 2014. Automatic detection of tulip breaking virus (TBV) in tulip fields using machine vision. Biosyst. Eng. 117, 35–42.
Qi, C., Sandroni, M., Westergaard, J.C., Sundmark, E.H.R., Bagge, M., Alexandersson, E., Gao, J., 2023. In-field classification of the asymptomatic biotrophic phase of potato late blight based on deep learning and proximal hyperspectral imaging. Comput. Electron. Agric. 205, 107585.
Qin, Z., Zhang, P., Wu, F., Li, X., 2021. Fcanet: Frequency channel attention networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 783–792.
Rai, N., Zhang, Y., Ram, B.G., Schumacher, L., Yellavajjala, R.K., Bajwa, S., Sun, X., 2023. Applications of deep learning in precision weed management: A review. Comput. Electron. Agric. 206, 107698.
Ramachandran, P., Zoph, B., Le, Q.V., 2017. Swish: a self-gated activation function. arXiv preprint arXiv:1710.05941 7, 5.
Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., Hughes, D.P., 2017. Deep learning for image-based cassava disease detection. Front. Plant Sci. 8, 1852.
Ravi, V., Acharya, V., Pham, T.D., 2022. Attention deep learning-based large-scale learning classifier for Cassava leaf disease classification. Expert. Syst. 39, e12862.
Sambasivam, G., Opiyo, G.D., 2021. A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. Egyptian Informatics Journal 22, 27–34.
Samborski, S.M., Tremblay, N., Fallon, E., 2009. Strategies to make use of plant sensors-based diagnostic information for nitrogen recommendations. Agron. J. 101, 800–816.
Santurkar, S., Tsipras, D., Ilyas, A., Madry, A., 2018. How does batch normalization help optimization? Advances in neural information processing systems 31.
Sethy, P.K., Barpanda, N.K., Rath, A.K., Behera, S.K., 2020. Deep feature based rice leaf disease identification using support vector machine. Comput. Electron. Agric. 175, 105527.
Stein, E.M., Shakarchi, R., 2011. Fourier analysis: an introduction. Princeton University Press.

Sun, R., Zhang, M., Yang, K., Liu, J., 2020. Data enhancement for plant disease classification using generated lesions. Appl. Sci. 10, 466.
Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. International Conference on Machine Learning. PMLR 6105–6114.
Tang, Y., Wang, X., Dellandrea, E., Chen, L., 2016. Weakly supervised learning of deformable part-based models for object detection via region proposals. IEEE Trans. Multimedia 19, 393–407.
Trockman, A., Kolter, J.Z., 2023. Patches are all you need? transactions on machine learning. Research.
Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., Nuflo, F., 1995. Modeling visual attention via selective tuning. Artif. Intell. 78, 507–545.
Ulyanov, D., Vedaldi, A., Lempitsky, V., 2016. Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022.
Wang, H., Wu, X., Huang, Z., Xing, E.P., 2020. High-frequency component helps explain the generalization of convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8684–8694.
Wang, H., Cheng, G., Li, Y., Cai, G., Sun, L., Wang, S., 2023. Person re-identification with IBN layer and channel attention module for indoor scenarios. In: Fourteenth International Conference on Graphics and Image Processing (ICGIP 2022). SPIE, pp. 139–151.
Wang, Y., Ostermann, J., Zhang, Y.-Q., 2002. Video processing and communications. Prentice hall Upper Saddle River, NJ.
Wang, W., Zhang, J., Wang, F., 2019. Attention bilinear pooling for fine-grained classification. Symmetry 11, 1033.
Wei, X.-S., 2023. Fine-grained image analysis: Modern approaches. Springer.
Weston, J., Chopra, S., Bordes, A., 2015. Memory networks, 3rd International Conference on Learning Representations, ICLR 2015.
Wu, W., Li, A.-D., He, X.-H., Ma, R., Liu, H.-B., Lv, J.-K., 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Comput. Electron. Agric. 144, 86–93.
Wu, H., Wiesner-Hanks, T., Stewart, E.L., DeChant, C., Kaczmar, N., Gore, M.A., Nelson, R.J., Lipson, H., 2019. Autonomous detection of plant disease symptoms directly from aerial imagery. The Plant Phenome Journal 2, 1–9.
Wydra, K., Verdier, V., 2002. Occurrence of cassava diseases in relation to environmental, agronomic and plant characteristics. Agr Ecosyst Environ 93, 211–226.
Yang, J., Li, C., Dai, X., Gao, J., 2022. Focal modulation networks. Adv. Neural Inf. Proces. Syst. 35, 4203–4217.
Yang, S., Xing, Z., Wang, H., Gao, X., Dong, X., Yao, Y., Zhang, R., Zhang, X., Li, S., Zhao, Y., 2023. Classification and localization of maize leaf spot disease based on weakly supervised learning. Front. Plant Sci. 14, 1128399.
Zárate-Chaves, C.A., Gómez de la Cruz, D., Verdier, V., López, C.E., Bernal, A., Szurek, B., 2021. Cassava diseases caused by Xanthomonas phaseoli pv. manihotis and Xanthomonas cassavae. Mol. Plant Pathol 22, 1520–1537.
Zhang, J., Bargal, S.A., Lin, Z., Brandt, J., Shen, X., Sclaroff, S., 2018. Top-down neural attention by excitation backprop. Int. J. Comput. vis. 126, 1084–1102.
Zhang, P., Dou, H., Yu, Y., Li, X., 2022b. Adaptive cross-domain learning for generalizable person re-identification. European Conference on Computer Vision. Springer 215–232.
Zhang, J., Qi, C., Mecha, P., Zuo, Y., Ben, Z., Liu, H., Chen, K., 2022a. Pseudo high-frequency boosts the generalization of a convolutional neural network for cassava disease detection. Plant Methods 18, 136.
Zhong, Y., Zhao, M., 2020. Research on deep learning in apple leaf disease recognition. Comput. Electron. Agric. 168, 105146.
Zhou, P., Feng, J., Ma, C., Xiong, C., Hoi, S.C.H., 2020. Towards theoretically understanding why sgd generalizes better than adam in deep learning. Adv. Neural Inf. Proces. Syst. 33, 21285–21296.
Zhu, W., Chen, H., Ciechanowska, I., Spaner, D., 2018. Application of infrared thermal imaging for the rapid diagnosis of crop disease. IFAC-PapersOnLine 51, 424–430.
Zou, X., Xiao, F., Yu, Z., Li, Y., Lee, Y.J., 2023. Delving deeper into anti-aliasing in convnets. Int. J. Comput. vis. 131, 67–81.