

DFN-PSAN: Multi-level deep information feature fusion extraction network for interpretable plant disease classification

Guowei Dai^a, Zhimin Tian^b, Jingchao Fan^{a,*}, C.K. Sunil^c, Christine Dewi^d

^a Agricultural Information Institute of CAAS, National Agriculture Science Data Center, Beijing 100081, China

^b Hebei University of Water Resources and Electric Engineering, Cangzhou, Hebei 061000, China

^c Department of Information Technology, National Institute of Technology Karnataka, Surathkal, 575025, India

^d Department of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia



ARTICLE INFO

Keywords:

Deep learning
Image processing
Feature fusion
Multilevel features
Pixel attention
Disease classification

ABSTRACT

Accurate identification of crop diseases is an effective way to promote the development of intelligent and modernized agricultural production, as well as to reduce the use of pesticides and improve crop yield and quality. Deep learning methods have achieved better performance in classifying input plant disease images. However, many plant disease datasets are often constructed from controlled scenarios, and these deep learning models may not perform well when tested in real-world agricultural environments, highlighting the challenges of transitioning to natural farm environments under the new demand paradigm of Agri 4.0. Based on the above reasons, this work proposes using a multi-level deep information feature fusion extraction network (DFN-PSAN) to achieve plant disease classification in natural field environments. DFN-PSAN adopts the YOLOv5 Backbone and Neck network as the base structure DFN and uses pyramidal squeezed attention (PSA) combined with multiple convolutional layers to design a novel classification network PSAN, which fuses and processes the multi-level depth information features output from DFN and highlights the critical regions of plant disease images with the help of pixel-level attention provided by PSA, thus realizing effective classification of multiple fine-grained plant diseases. The proposed DFN-PSAN was trained and tested on three plant disease datasets. The average accuracy and F1-score exceeded 95.27%. The PSA attention mechanism saved 26% of model parameters, achieving a competitive performance among existing related methods. In addition, this work effectively enhances the transparency of the features of the model attention to plant diseases through t-SNE with SHAP interpretable methods.

1. Introduction

As the cornerstone of national development, agriculture is vital to the national economy. Currently, the global food supply is tight, and the yield and quality of crops, as a strategic and fundamental core industry of the country, are often seriously affected by the climatic environment, pests and diseases, and infestation by various pathogens (Jayagopal et al., 2022; Legrand, 2023). Thousands of common crop diseases have been identified, and traditional manual identification methods are time-consuming and ineffective. Therefore, accurate and rapid identification of crop diseases is essential to take timely relief measures and reduce yield and quality losses.

As of now, many studies have been conducted on crop disease phenomes, mainly focusing on the identification and classification of

diseases. Deep learning methods produce higher recognition and classification accuracy than traditional machine vision techniques, as they have a more powerful feature representation and automatic feature learning. In addition, the deep learning end-to-end network structure has high recognition speed and accuracy (Thakur et al., 2022). Classical neural networks are used in plant disease classification studies, and the popular AlexNet, GoogleNet, VGGNet, ResNet, and DenseNet models are often used as the basis of research to achieve performance improvements by combining other methods (Wani et al., 2022). Dogra et al. (2023) used two-stage fine-tuning to complete the weight migration of VGG19 and achieved 93.0 % recognition accuracy for rice leaf diseases. Yu et al. (2023a), based on the pre-trained model of the ImageNet dataset, migrated its convolutional layer weights to the convolutional layer of the ResNet18 model and thus proposed the TRNet18 model; the OTSU

* Corresponding author at: No. 12, Zhongguancun South Street, Haidian District, Beijing, China.

E-mail addresses: dgwstyle@foxmail.com (G. Dai), fanjingchao@caas.cn (J. Fan).

method was used to remove the sample background of the classification dataset, and the diseased leaf images were expanded by applying data augmentation methods after regional calibration; in the classical neural network model AlexNet, the ResNet18, and ResNet50 had the most robust performance in comparison with an average recognition rate of 99.53 %. Inspired by deep convolutional neural networks for automatic feature extraction, Reddy, Varma, and Davuluri (2023) proposed to extract color and texture features of plant leaf images using ResNet50, followed by MRDOA-inspired bio-optimization method to select the optimal features and designed a simple classification convolutional neural network DLCNN for classifying the selected plant leaf disease features in PlantVillage and rice datasets with a classification accuracy of 99.73 % and 68.99 %, respectively. Pavithra, Kalpana, and Vigneswaran (2023) proposed a new deep learning-based automated plant disease classification model, DL-APDDC, using U2Net to remove background regions of potato leaves and citrus fruits, followed by the Adam-optimized SqueezeNet model as a feature extractor and an extreme gradient boosting (XGBoost) classifier to classify the diseases.

Custom neural networks or improved classical neural networks for plant disease classification are studied by improving the network structure, adding attention mechanisms, stacking more convolutional units, and introducing a visual Transformer to enhance the model performance. Alirezazadeh et al. (2023) verified the effectiveness of embedding the CBAM attention mechanism in the tails of well-known pre-trained CNN models (i.e., EfficientNetB0, MobileNetV2, ResNet50, InceptionV3, and VGG19), and the experimental results highlighted the ability of CBAM to help deep learning models overcome the problem of lack of sufficient training data. Hassan and Maji (2022) proposed a novel lightweight CNN architecture based on Inception and ResNet, which reduced the parameters by 70 % and obtained a performance accuracy of 99.39 % on a rice disease dataset. Introducing SE, CBAM, and ECA attention modules into the convolutional model enhanced the CNN model's focus on plant diseases and reduced the impact of complex background on disease identification (Fang et al., 2023; Ni et al., 2023; Zhu et al., 2023). Yu, Xie, and Huang (2023) designed a Transformer-based neural network by introducing a visual Transformer in the Inception model and using migration learning and fine-tuning methods to avoid overfitting problems. Experimental results showed that 99.94 % and 99.22 % accuracy was achieved on PlantVillage and Ibean, respectively. Better than existing models. Kaya and Gürsoy (2023) proposed to use RGB and segmented images of plant diseases as input to the DenseNet121 model and to achieve information fusion by combining the complementary knowledge of features from multiple sources. Combinatorial transformation of multilevel features can enrich feature representation, and the proposed adaptive channel, spatial and pixel attention multilevel feature fusion network (MFFN), based on ResNet50 by stacking multiple MFFN layers to combine the obtained multilevel features for classification, achieved 99.88 % test accuracy (Sunil et al., 2023). The above study applied deep learning to plant disease classification with good results.

Focusing on the research of detecting the distribution of spots and lesions in different forms of plant diseases, Qi et al. (2022) proposed to add a squeeze-and-excitation (SE) attention module to the YOLOv5 model to achieve the extraction of key features with an accuracy of 91.07 % for effective detection of tomato virus regions. Li et al. (2022) improved YOLOv5 by integrating the improved Inception module with the Feature Pyramid Network (FPN) to obtain multi-scale disease features and achieved 93.1 % mAP for vegetable disease detection. Sending all disease samples to the model for detection can consume much time; therefore, filtering healthy leaf images or severely damaged leaf images using the Xception to reduce processing time, followed by Faster-RCNN to locate and detect plant disease spots, showed that a high detection performance was achieved at the same time with better detection speed (Khan et al., 2022). The YOLOv5 has been adopted by a large number of plant disease detection researchers due to its faster recognition speed and feature extraction capability (Chen et al., 2023; Li et al., 2023), and

the end-to-end detection model outputs the detection results directly, but this requires a large amount of data for crucial feature annotation. It consumes more human, material, and financial resources than labeled classification data. Secondly, for smart plant protection agriculture scenarios, determining key disease characteristics is of little value to farmers. From the precision agriculture perspective, disease severity estimation helps address pesticide misuse and abuse. Although single diseased leaf spots can be quantified by detecting and locating them, it brings more computational costs. In contrast, disease classification statistics by dividing certain farm areas into individual leaf units can be equally helpful for quantifying diseases. Therefore, disease classification research models have a critical role in agricultural production.

Based on the above considerations, we designed a novel lightweight deep learning model based on the YOLOv5 network to identify plant diseases. This paper improves the original YOLOv5 network by keeping the feature extraction network Backbone and the feature fusion network Neck, removing the Head structure, and designing a novel PSAN classification network structure for plant disease classification; Pyramid Squeeze Attention (PSA) is used for PSAN, which allows the network to focus on the features it features that need attention and ignore the unimportant features. The model is trained on three plant disease datasets close to real agricultural scenarios, and the performance is evaluated. The main contributions of this paper are summarized as follows:

(1) The DFN structure of the DFN-PSAN model aims to obtain information on plant disease characteristics at different scales through image feature fusion techniques, which migrate the advantages of the target detection network structure to the classification network.

(2) The PSAN structure of the DFN-PSAN model makes full use of the DFN structure to provide rich information on important semantic features, and the embedded PSA attention mechanism can reinforce important information to suppress non-important information and improve neural network performance.

(3) The two-stage weather data augmentation technique is used for plant disease datasets in three real agricultural scenarios, which improves the depth model generalization performance and suppresses model overfitting.

(4) The t-SNE method is used to interpret the feature layer data of the DFN-PSAN model by visualizing the two-dimensional clustering distribution of the feature data to show the transparency of the model.

(5) The SHAP interpretable AI (XAI) visualization method explains whether DFN-PSAN can correctly focus on plant disease features or pattern information.

The rest of this study is organized as follows: Section 2 describes three plant disease datasets in detail, followed by data pre-processing and augmentation, experimental working conditions, the PSA attention mechanism, the principle and workflow of the proposed DFN-PSAN model, and a detailed description of the essential hyperparameters of the model. Section 3 analyzes the results of DFN-PSAN disease identification and understands the model attention feature patterns through an interpretable approach. Section 4 discusses the characteristics of the work in this paper relative to other studies. Section 5 presents the conclusions and future work of the paper.

2. Materials and methods

2.1. Acquisition of datasets

The images in the PlantVillage open-source plant disease dataset are taken with a uniform background, which does not accurately reflect the complex plant disease symptoms found in natural field environments; therefore, three plant disease datasets constructed from actual scene collections were selected as validation objects for this study, and the detailed data of the various types of samples in the datasets are shown in Table 1.

Katra-Twelve is a public dataset of leaf images provided by the University of Shri Mata Vaishno Devi in Katra, which consists of healthy

Table 1

Detailed descriptions of the various types of samples comprising the three datasets.

Classes	Numbers diseased	Numbers healthy
Katra-Twelve		
Mango	265	170
AlstoniaScholaris	254	179
Jamun	345	279
PongamiaPinnata	276	322
Pomegranate	272	287
Chinar	120	103
Gauva	142	277
Arjun	232	220
Jatropha	124	133
Lemon	77	159
Bael	118	—
Basil	—	148
BARI-Sunflower		
Downy mildew	470	—
Leaf scars	509	—
Gray mold	398	—
Fresh leaf	—	512
FGVC8		
complex	1602	—
frog_eye_leaf_spot	3181	—
frog_eye_leaf_spot complex	165	—
healthy	—	4624
powdery_mildew	1184	—
powdery_mildew complex	87	—
rust	1860	—
rust complex	97	—
rust_frog_eye_leaf_spot	120	—
scab	4826	—
scab_frog_eye_leaf_spot	686	—
scab_frog_eye_leaf_spot complex	200	—

and diseased leaves. The dataset contains 12 economically and environmentally beneficial plants with 22 leaf types, consisting of 4503 images, including 2278 images of healthy leaves and 2225 images of diseased leaves (Chouhan et al., 2019). The images were acquired at a resolution of 6000×4000 pixels in JPG format, as shown in Fig. 1 dataset (a), with the upper and lower images indicating healthy and diseased leaves, respectively, labeled 0–9 for Mango, Alstonia Scholaris, Jamun, Pongamia pinnata, Pomegranate, Chinar, Guava, Arjun, Jatropha, and Lemon. In addition, 10 is the diseased leaf of Bael, and 11 is the healthy leaf of Basil.

The dataset BARI-Sunflower was constructed from the demonstration farm collection of Bangladesh Agricultural Research Institute (BARI), Gazipur, which contains 467 raw images of delicate leaves and infected sunflower leaves and flowers (Sara et al., 2022). To accommodate the condition that a large amount of data is required to train the deep learning model, spatial location data augmentation was accomplished by random rotation, scaling, cropping, shifting, noise, and blurring; on the other hand, color enhancement was accomplished by changing brightness, contrast, saturation, and hue. The enhanced samples were 470 instances of Downy mildew, 509 instances of Leaf scars, 398 instances of Gray mold, and 515 instances of Fresh leaf, respectively, represented by 0–3 of the dataset (b) in Fig. 1, with an image size of 512×512 pixels and format of JPG.

The dataset FGVC8 was constructed by Cornell Initiative for Digital Agriculture (CIDA), which significantly increased the number of apple leaf disease images based on FGVC7 and added more fine-grained disease categories, with a total of 18,632 images with 12 categories containing six significant categories and an image size of 4000×2672 pixels. The main disease categories are shown in Fig. 1 dataset (c), consisting of leaf images taken at different maturity stages and at different times of the day at different focal length camera settings, with the labels 0–5 indicating Complex, Frog eye leaf spot (FLS), Healthy, Powdery mildew, Rust, and Scab, respectively.

2.2. Data pre-processing

Data pre-processing is divided into two stages in this study, the first stage is the preliminary basic image pre-processing, and the next stage is the combination of techniques with weather data augmentation to complete the plant leaf data augmentation for agriculture-specific application scenarios, both of which work together to achieve a data augmentation pre-processing method with streaming.

The preliminary image preprocessing is based on the OpenCV image read-in (imread) method to read the image in BGR format by default, and then the format conversion (cvtColor) method is needed to convert the BGR to RGB format image. When the format conversion is satisfied, adaptive scaling is used to prevent the loss of image information by scaling each image size to 256×256 pixels. The Gaussian filter (GaussianBlur) method is then used to smooth the image by replacing each pixel value with the average of the surrounding pixels to reduce the image noise and detail with Gaussian kernel size 3. The dataset contains plant canopy images that are all in color, and in order to retain the color image detail while effectively removing the noise from the image, the non-local mean noise reduction (fastNlMeansDenoisingColored) method is used to preserve the object texture as much as possible, with parameters h and hColor both being 3. Finally, it is usually necessary to adjust the image brightness and contrast properties in computer vision tasks, using the direct manipulation of the image matrix with multiplication factors and offsets (convertScaleAbs) method to control the brightness and contrast, which needs to be adjusted with parameters $\alpha = 1.00$ and $\beta = 0$.

A weather data augmentation technique was used in the second data preprocessing phase to simulate environmental changes under different weather conditions. We used a solar illumination transformation (RandomSunFlare) technique to simulate natural illumination and light adjustment effects, where the appearance region of solar flares is determined by four parameters (x_{\min} , y_{\min} , x_{\max} , y_{\max}). In this paper, we choose the image's upper right corner (0.9, 0, 1, 0.5) and upper left corner (0.0, 0.0, 1.0, 0.1) as the source of illumination and set the aperture radius parameter to 300. To simulate the effect of rain on the image, we used the raindrop transformation (RandomRain) technique to add a random raindrop effect. We chose a raindrop size of 1.0, set the type to drizzle, and set the brightness coefficient of 0.6. Natural agricultural environments have close plant spacing and are susceptible to shadows from leaf shading, and the shadow transformation (RandomShadow) technique we used eliminates this effect to some extent. In this study, we randomized the area where shadows appear and set the number of shadows to float between 1 and 5, while the side parameter of the shadow polygon was set to 6. The fog occurs due to the high temperature and humidity differences in natural agricultural environments. To simulate this situation, we used the fog transformation (RandomFog) technique to add fog randomly to different locations of the image and to blur the background. We chose a fog intensity (fog coef) and fog circle transparency (alpha coef) between 0.25 and 0.8 and set the parameter value of fog to 0.3. The four weather data augmentation methods described above can be executed in random combinations with an execution probability 0.6 for each method.

The image processing was performed by VSCode and Python 3.10. Pytorch 1.13.1 + cu117 was used with the OpenCV library and accelerated by the GPU; Table 2 describes the detailed parameters of the test environment and dataset. Due to the imbalance of apple leaf disease categories in dataset FGVC8, six data augmentations were performed per image for categories with sample volumes below 500. Fig. 2 compares the augmented data. The Katra-Twelve and BARI-Sunflower datasets performed only two data augmentations per image. Taking the Downy mildew sample in dataset BARI-Sunflower as an example, the fundamental image preprocessing was executed first, followed by the execution of weather data augmentation to complete the simulation of data collected under different weather conditions, and the execution effect of the two-stage image processing is shown in Fig. 3. After preprocessing,



Fig. 1. Fine-grained plant disease images of natural agricultural scenes. (a) Katra-Twelve; (b) BARI-Sunflower; (c) FGVC8.

Table 2
Dataset description and test system environment configuration.

Parameter	Configuration
Dataset description	
Katra-Twelve	9006
BARI-Sunflower	3784
FGVC8	21,977
dimension	256 × 256
formats	JPG
System environment	
Operating System	Windows 11 Professional Workstation Edition
CPU	Intel(R) Core (TM) i9-13900 K
RAM	64.00 GB
GPU	NVIDIA GeForce RTX 3090

the three new datasets formed were randomly divided into three parts according to the 8:1:1 division, including the training, validation, and test sets.

2.3. Pyramid Squeeze attention

The Pyramid Squeeze Attention (PSA) module is crucial for computer vision tasks. It enriches the feature space by exploring channel feature information at multiple scales, efficiently extracting multi-scale spatial

information at a finer granularity level, and adaptively recalibrating the attention weights across channels (Zhang et al., 2022). The PSAN network uses the PSA module for multi-scale feature fusion. In the following, we will detail the specific flow of the PSA attention module, as shown in Fig. 4. Given an input feature map X, it has C channels. We first divide the input X into S groups and calculate the weights of each channel in each group. In this study, we set S to be 4. The number of channels in the output of each convolution group is consistent and is obtained by computing c/s . Thus, we obtain a multi-scale feature representation that can be expressed as:

$$F_i = \text{Conv2d}(k_i \times k_i, G_i)(X) \quad (i = 0, 1, \dots, S - 1) \quad (1)$$

Where Conv2d denotes the two-dimensional convolution operation for extracting the spatial information in the feature map, the convolution kernel size is determined by the formula $2 \times i + 3$, where i denote the group's index. k_i denotes the convolution kernel size of the i th group convolution, and according to the formula $2 \times i + 3$ we can determine the convolution kernel size of each group convolution. G_i denotes the parameter of the i th group convolution, which is used to adjust the feature extraction ability of that group convolution. In this study, we set the S-group convolution kernel size to $k=\{3,5,7,9\}$ and control the number of groups in each group convolution by adjusting the parameter $G=\{1,4,8,16\}$. We can obtain the whole multi-scale feature map by group convolution and feature fusion. Specifically, we join the feature

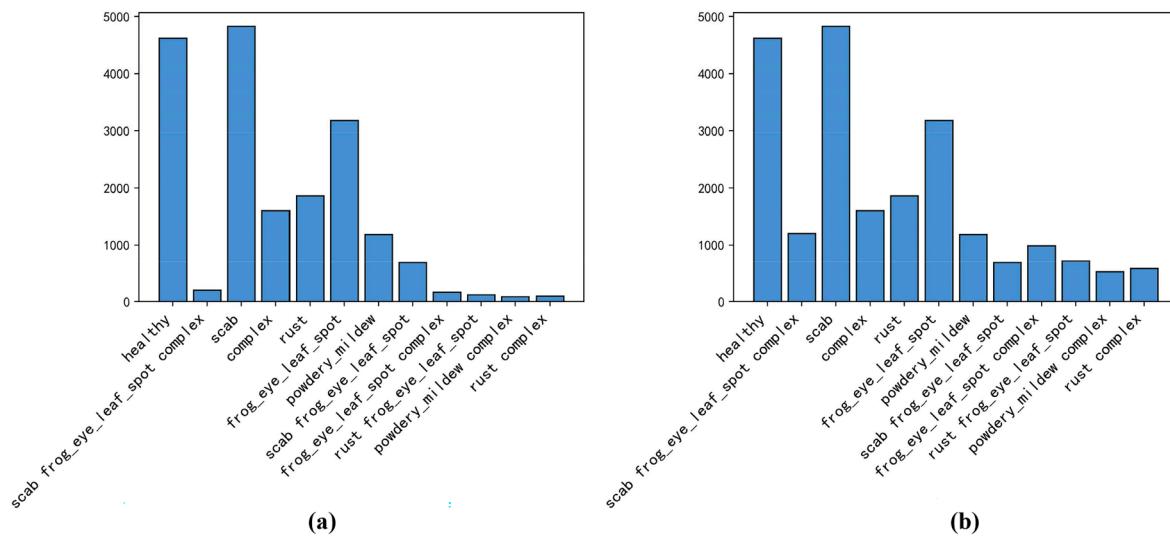


Fig. 2. Comparison of various sample sizes after data augmentation performed by FGVC8. (a) FGVC8; (b) FGVC8 after data augmentation.

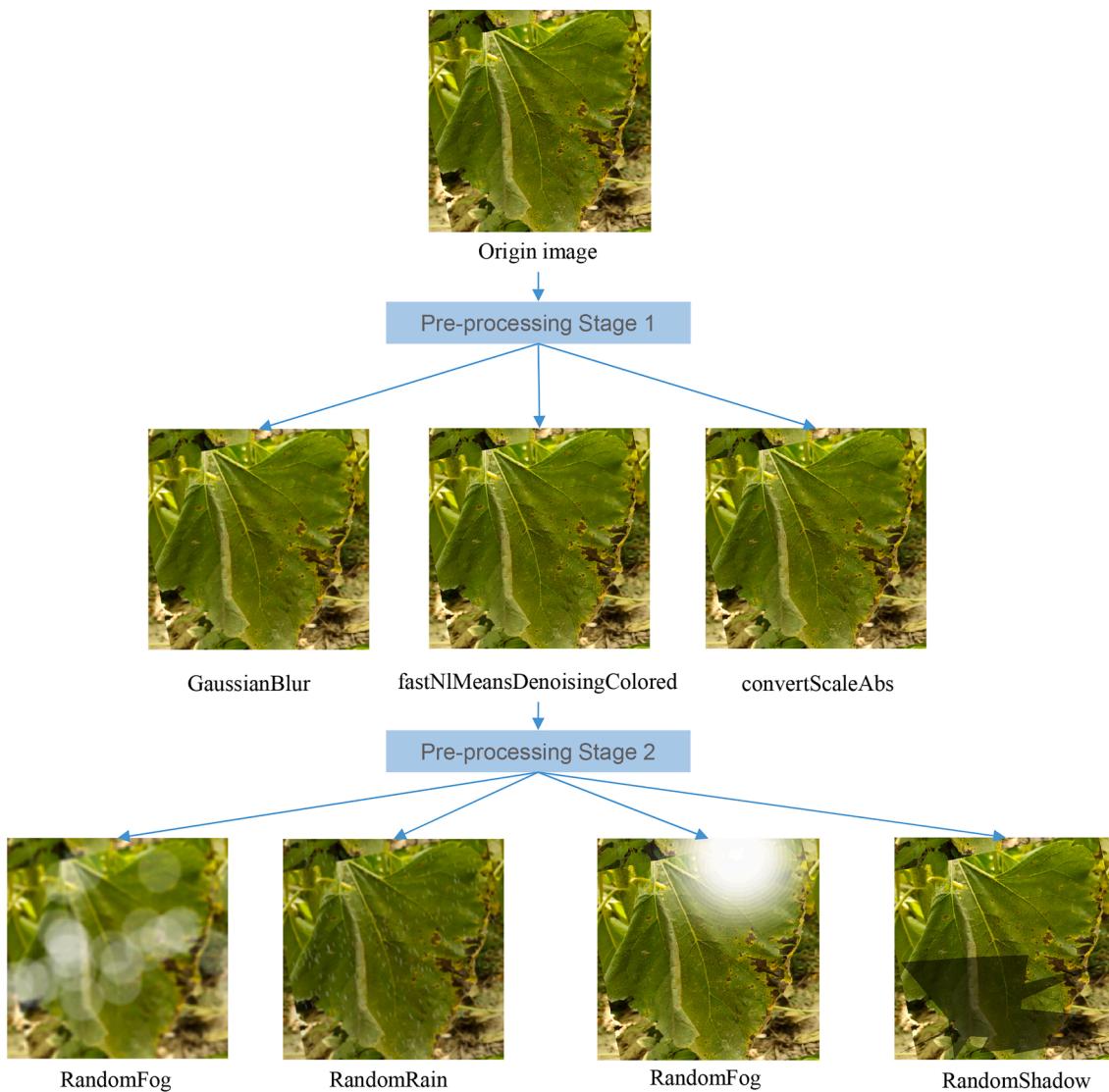


Fig. 3. Image pre-processing and data augmentation method with two stages.

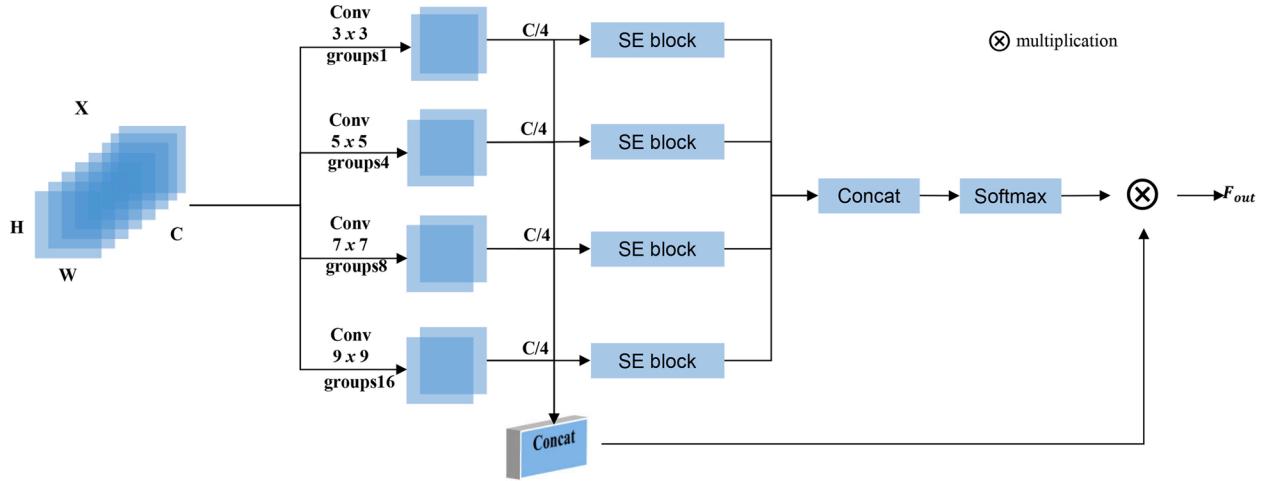


Fig. 4. Detailed description of the PSA attention module structure.

maps of multiple branches to obtain the multi-scale feature map representation shown below:

$$F = \text{Concat}([F_0, F_1, \dots, F_{S-1}]) \quad (2)$$

Among them, we use the feature map join operation *Concat* for joining multiple channels of multi-scale feature maps $F_i (i = 0, 1, \dots, S-1)$ together. The dimensions of these feature maps are $F \in R^{H \times W \times C}$, where H, W, and C denote the height, width, and number of channels of the feature maps, respectively. Next, we introduce the channel attention mechanism *SEWeight* to calculate the channel weights for each group. The learned channel attention weights are multiplied with the original feature map F_i at the element level to obtain the feature map with reassigned weights. It can be expressed as:

$$\text{SEW} = \text{Concat}(\text{SEWeight}(F_i)), (i = 0, 1, \dots, S-1) \quad (3)$$

Where *Concat* is used to connect the attention weights of different channels to form a multi-scale attention weight vector *SEW*. Next, we recalibrate this multi-scale attention weight vector with the *Softmax* function to ensure that the sum of the weights of each channel is equal to 1. This makes the multi-scale channel attention weights more expressive and adaptive. At the same time, we perform the multiplication operation (\odot) on channels with the: multi-scale attention weights *SEW* and the multi-scale feature map F, i.e., the element-level multiplication operation. This operation can effectively integrate the different sizes of perceptual fields and rich multi-scale spatial information. Eventually, we obtain the output multi-scale feature map, which can be expressed as:

$$F_{out} = F \odot \text{Softmax}(\text{SEW}) \quad (4)$$

2.4. Proposed DFN-PSAN architecture

The DFN-PSAN model consists of feature extraction and classification, and YOLOv5 takes the main job of feature extraction. A popular image classification algorithm in the YOLO family is You Only Look Once Version 5 (YOLOv5). This machine-learning technique can be used for real-time applications due to its improved computational accuracy and speed. YOLOv5 is the 5th generation algorithm of the YOLO series, among which YOLOv5n has the smallest weight value, small detection network depth, and fast detection speed. However, the method has a relatively low extraction ability of deep image features. It is prone to miss detection and classification errors in various types of classification detection, which cannot meet the accuracy requirements of plant disease recognition. In this paper, YOLOv5n is improved and enhanced by reconstructing the convolutional neural network of the classification

layer, which improves the feature extraction and feature fusion ability of recognition and the convergence speed.

The YOLOv5 architecture consists of three main parts, which are composed of a backbone network module (Backbone), a feature fusion module (Neck), and a prediction module (Head). In the proposed DFN-PSAN model, Backbone and Neck form the DFN and are used to obtain deep fusion feature information. The classification layer convolutional neural network PSAN replaces the Head designed to improve the classification performance of plant diseases, and its network structure is shown in Fig. 5.

The Backbone is located after the input image and is used for feature extraction and processing the required feature output at each stage. the Backbone network uses the CSPDarkNet network structure, replacing the previous version of the Focus structure with a 6×6 convolutional layer (Conv) in the first layer, which is more efficient for existing GPUs and optimized algorithms. C3 is used to extract features from the feature map. Compared to other large-scale CNNs, the C3 structure reduces the repetition of information gradients in the optimization part of the neural network process, which represents one of the critical parts of the overall YOLOv5 network structure, since its parameters occupy the most significant part of the overall network parametric count. The idea of scalability of the neural network is also reflected when changing the width and depth of the C3 part of the network. The spatial pyramid pooling (SPP) module is used to increase the perceptual field of the network and realize the extraction of local and global features. The input feature map (FM) is operated by maximum pooling with the same step length but different sizes to obtain local and global features. Then the local and global features are fused using the concatenation (Concat) operation to enrich the expression capability of the feature map and obtain additional network features of different scales.

The Neck combines the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures. As the depth of the network increases, the feature map changes to a smaller size and more profound depth. The large-sized feature maps at the bottom contain more accurate localization information, while the small-sized ones at the top contain more abstract semantic information. The use of FPN to convey vital semantic information by upsampling from the top down and PAN to convey vital localization information by downsampling from the bottom up complement each other to improve the feature fusion capability of the algorithm. The features are a combined aggregation of extracted features from different layers to further improve the network's performance and, thus, the ability to detect objects at different scales. The YOLOv5 has 30 different training hyperparameters, and these procedures significantly improve the effectiveness of the YOLOv5 system and enable its use in real time.

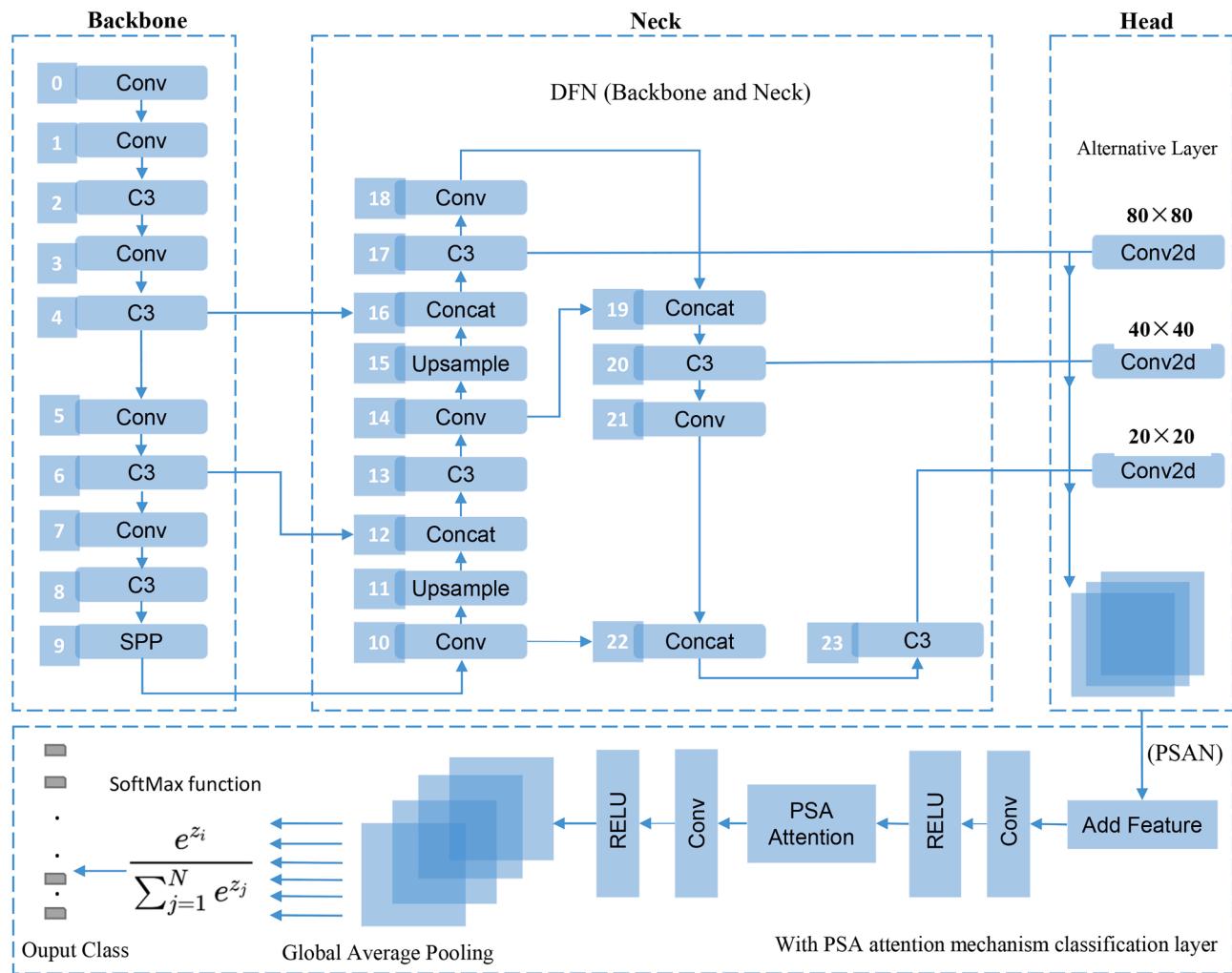


Fig. 5. Technology roadmap for the DFN-PSAN model architecture.

The DFN-PSAN model reconstructs the classification layer of YOLOv5, i.e., the Head of the original YOLOv5 is replaced by the PSAN classification layer in Fig. 5. The new convolutional neural network structure is used for feature classification, and the Neck structure of YOLOv5 provides in-depth information fusion features. Convolutional feature networks are essential for classification, and the proposed model is designed with several convolutional layers that have been taught to identify the distinctive plant leaf patterns. The convolutional layers define how to perform convolutional operations across the entire set of outstanding plant disease features using various filters. The PSA attention mechanism is incorporated into the convolutional neural network to improve the model's ability to perceive helpful information while suppressing unwanted noise. After that, the salient information extracted by the attention mechanism is transmitted to the global average pooling (GAP) layer to more adequately extract the classification target information. The final Softmax layer assigns a corresponding prediction probability to each plant leaf category. The whole process consists of a deep neural network updating the trainable model parameters by comparing the difference between the expected and current models for each gradient update of a batch of data. The details of constructing the new PSA attention mechanism used as a PSAN model for the classification layer convolutional neural network are shown in Table 3.

2.5. Model hyperparameters

In this study, all hyperparameters are shown in [Table 4](#). The setting

of hyperparameters can impact the model performance and recognition accuracy. For fast convergence of the model, we randomly sampled and calculated the mean and standard deviation of the images for all samples of the three disease datasets. We normalized the calculated values for the images in the training phase. The optimization function prioritizes the use of Adam to get through the previous values quickly, followed by tuning the model using SGD. The YOLOv5 classification module has a generic approach for image augmentation, and these hyperparameters are selected based on ImageNet. However, the image augmentation hyperparameters are reacquired using an automatic hyperparameter search to better match the plant disease scenario. The EarlyStopping is integrated into DFN-PSAN to prevent model overfitting, and the DFN-PSAN channel is set to a suitable minimum value with the model parameters. In the data pre-processing stage, weather data augmentation techniques can mitigate overfitting. To further enhance the model's generalization ability, Label Smoothing is added to the cross-entropy loss function, and Label Smoothing reduces the risk of model overfitting by adding noise to the output to constrain the model (Gao et al., 2022). The cross-entropy loss function with Label Smoothing added can be expressed as Eqs. (5), (6), where N is the total number of categories, i is one of the categories, y_i is the prediction result of i , p_i is the confidence score of the network output, and ϵ is the artificially set hyperparameter.

$$X_{\text{Loss}} = - \sum_{i=1}^N y_i \log(p_i) \quad (5)$$

Table 3

Detailed parameters of PSAN model of classification layer convolutional neural network.

Modules	Types	Kernel size	Stride	Padding	Output size
Conv2d-1	Convolution	(3,3)	1	1	(64, 256, 256)
ReLU-2	Activation Function	—	—	—	(64, 256, 256)
Conv2d-3	Convolution	(3,3)	1	1	(128, 256, 256)
ReLU-4	Activation Function	—	—	—	(128, 256, 256)
Conv2d-5	Convolution	(3,3)	1	1	(256, 256, 256)
ReLU-6	Activation Function	—	—	—	(256, 256, 256)
Conv2d-7	Convolution	(3,3)	1	1	(64, 256, 256)
Conv2d-8	Convolution	(5,5)	1	2	(64, 256, 256)
Conv2d-9	Convolution	(7,7)	1	3	(64, 256, 256)
Conv2d-10	Convolution	(9,9)	1	4	(64, 256, 256)
AdaptiveAvgPool2d-11	Average Pooling	—	—	—	(64, 1, 1)
Conv2d-12	Convolution	(1,1)	1	1	(4, 1, 1)
ReLU-13	Activation Function	—	—	—	(4, 1, 1)
Conv2d-14	Convolution	(1,1)	1	1	(64, 1, 1)
Sigmoid-15	Activation Function	—	—	—	(64, 1, 1)
SEWeightModule-16	Channel Attention	—	—	—	(64, 1, 1)
AdaptiveAvgPool2d-17	Average Pooling	—	—	—	(64, 1, 1)
Conv2d-18	Convolution	(1,1)	1	1	(4, 1, 1)
ReLU-19	Activation Function	—	—	—	(4, 1, 1)
Conv2d-20	Convolution	(1,1)	1	1	(64, 1, 1)
Sigmoid-21	Activation Function	—	—	—	(64, 1, 1)
SEWeightModule-22	Channel Attention	—	—	—	(64, 1, 1)

$$y_i = \begin{cases} 1 - \epsilon, & i \text{ is target} \\ \epsilon/N, & i \text{ is not target} \end{cases} \quad (6)$$

3. Experimental results and analysis

3.1. Model results analysis

This section verifies the performance of the DFN-PSAN model on the three datasets. As shown in Fig. 6, DFN-PSAN performs the best on the Katra-Twelve dataset, the worst on the FGVC8 dataset, and the performance for the BARI-Sunflower dataset is in between. The overall model performance corresponds to the sequentially increasing complexity of the disease images for the three datasets Katra-Twelve, BARI-Sunflower & FGVC8. In a comprehensive comparison, the average accuracy and F1 scores of DFN-PSAN exceeded 95.27 % for all three datasets.

In addition, to verify the performance stability of the model on the three datasets, Using the idea of k-fold cross-validation, we processed

the three datasets sequentially, with each dataset randomly divided into four parts. In each division, we use 30 % of the data as a test set, while the remaining 70 % is combined with the remaining three parts to form a new training set. This ensures that each part is used as a test set in one of the divisions. The DFN-PSAN model is trained on the training set, validated on the test set, and the validation results are recorded. Table 5 shows the results of the 4-fold cross-validation. The average accuracies on Katra-Twelve, BARI-Sunflower, and FGVC8 were 98.37 %, 94.23 %, and 93.24 %, respectively, and the accuracy fluctuated no more than 3 % in the cross-validation. The more significant variation in accuracy is on the FGVC8 dataset, which may be due to the random partitioning of this dataset into the test set with disease images that are difficult to discriminate on visual features, leading to errors in DFN-PSAN identification during testing, as confirmed by the Table 5 dataset FGVC8 on the 1-flop vs. 2-flop cross-test of testing.

In this study, to demonstrate the usefulness of the weather data augmentation technique in the second stage of data preprocessing, we also validated the performance stability of the model applying the weather data augmentation technique in the three data sets based on the k-fold cross-validation idea. Following the method used in this study to validate the model performance stability in the three datasets, each was randomly divided into three parts, and the experimental results are shown in Table 6. With the second stage of weather data augmentation, the average accuracy of DFN-PSA on three datasets, Katra-Twelve, BARI-Sunflower, and FGVC8, improved by 0.69 %, 2.99 %, and 1.09 %, respectively, and primarily worked better on BARI-Sunflower and FGVC8. Through data analysis, the Katra-Twelve dataset has a single image background, thus allowing direct visual focus on the subject or lesion area of the diseased leaves. In contrast, the BARI-Sunflower dataset has a smaller original sample size and was taken under various lighting, angle, surface, and noise conditions, with more significant data quality differences and visual feature complexity issues, which is the main reason for the less effective data augmentation than the former. Compared with the other two types of data samples, FGVC8 has many original data samples with more complex disease identification types. However, the data generalization capability brought by the two-stage data augmentation eases the difficulty of extracting features from the DFN-PSAN model under complex feature patterns. It effectively makes up for the ability of the model to expand multi-dimensional feature extraction. Therefore, the two-stage weather data augmentation technique has better effectiveness in the case of complex visual feature patterns with a large volume of sample data.

3.2. t-SNE Interpretability analysis

In this section, for the multidimensional and nonlinear characteristics of plant disease feature data, the t-distributed stochastic neighbor embedding (t-SNE) algorithm will be used to visualize the feature layer data with two-dimensional t-SNE and then improve the transparency and interpretability of the DFN-PSAN plant disease identification model. DFN-PSAN designs PSAN with a novel incorporated attention mechanism as a new classification network. PSAN has unfolded the feature layer and attention layer compared to the original classification network of YOLOv5, and this design can obtain fine-grained attention power in multidimensional data. To explain the ability of this part of the model to focus on disease features, we extracted data features from two-dimensional visualizations under three test sets. Strictly speaking, the models involved in the comparison are divided into DFN-PSAN and general classification network DFN.

Fig. 7 shows the better clustering visualization of the nonlinear dimensionality reduction technique (t-SNE) in highlighting the feature patterns. For the simple background KatraTwelve dataset, Fig. 7 (b) shows more clusters than Fig. 7 (a) and better aggregation of the same cluster samples. Both Fig. 7 (c) and Fig. 7 (d) can be divided into 4 clusters, but the sample points in Fig. 7 (d) have almost no overlap. The samples in different clusters are far away from each other, which

Table 4
Optimized hyperparameter values of the DFN-PSAN model.

Experimental Settings	Hyperparameter	Optimized value
Training Settings	batch size	64
	optimizer	SGD, Adam
	momentum	0.9000
	weight decay	0.0005
	learning rate	0.0100
	loss function	CrossEntropyLoss
	epochs	30
Katra-Twelve Normalize	mean	[0.23567992, 0.2659049, 0.24378718]
	std	[0.12026429, 0.1425962, 0.11200316]
BARI-Sunflower Normalize	mean	[0.4516689, 0.48817652, 0.26045147]
	std	[0.22680794, 0.2133477, 0.18061319]
FGVC8 Normalize	mean	[0.40568268, 0.51418954, 0.3238448]
	std	[0.2018306, 0.18794753, 0.18870047]
CrossEntropyLoss	label_smoothing	0.1
ModelSize	depth_multiple	0.33
	width_multiple	0.25
EarlyStopping	patience	6
ClassifyAlbumentations	random resized crop	scale = [0.08, 1.0]; ratio = [0.75, 1.12]
	horizontal flip	0.5
	vertical flip	0.1
	color jitter	0.2

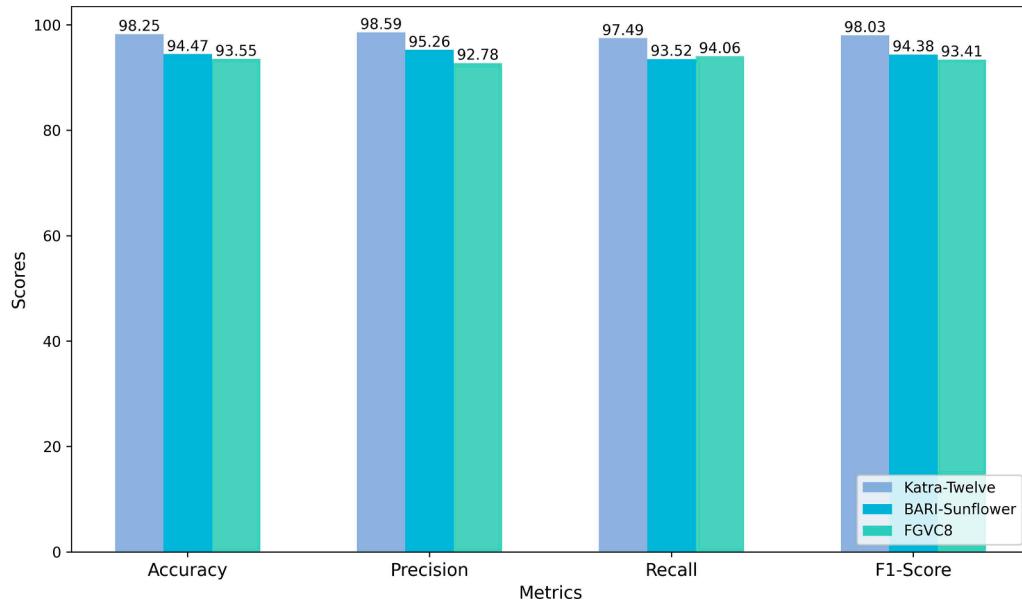


Fig. 6. Classification test performance of the DFN-PSAN model on three datasets.

Table 5
DFN-PSAN test results based on k-fold cross-validation.

No of Fold	Accuracy (%)		
	Katra-Twelve	BARI-Sunflower	FGVC8
1-fold	97.89	93.25	91.98
2-fold	99.06	93.68	94.53
3-fold	98.11	95.21	93.26
4-fold	98.44	94.79	93.19
Average	98.37	94.23	93.24

preserves the original data's local structural features and shows the ability of DFN-PSAN to identify plant diseases in real agricultural scenarios. On the FGVC8 dataset, both Fig. 7 (e) and Fig. 7 (f) have overlapping color blocks, consistent with the feature that FGVC8 labels the same disease sample as multiple diseases when defining plant disease data samples. In addition, the disease data samples labeled by FGVC8 are divided into six parts. Although both Fig. 7 (e) and Fig. 7 (f) have six

Table 6
Experimental results of DFN-PSA model augmentation in two-stage pre-processed data.

Method	Accuracy (%)			
	Katra-Twelve	BARI-Sunflower	FGVC8	—
With Pre-processing	98.11	93.87	92.99	—
	98.64	94.35	93.73	—
	99.06	95.12	92.59	—
	98.54	94.44	93.10	average
Without Pre-processing	97.98	92.12	91.85	—
	98.03	90.79	92.17	—
	97.54	91.46	92.03	—
	97.85	91.45	92.01	average

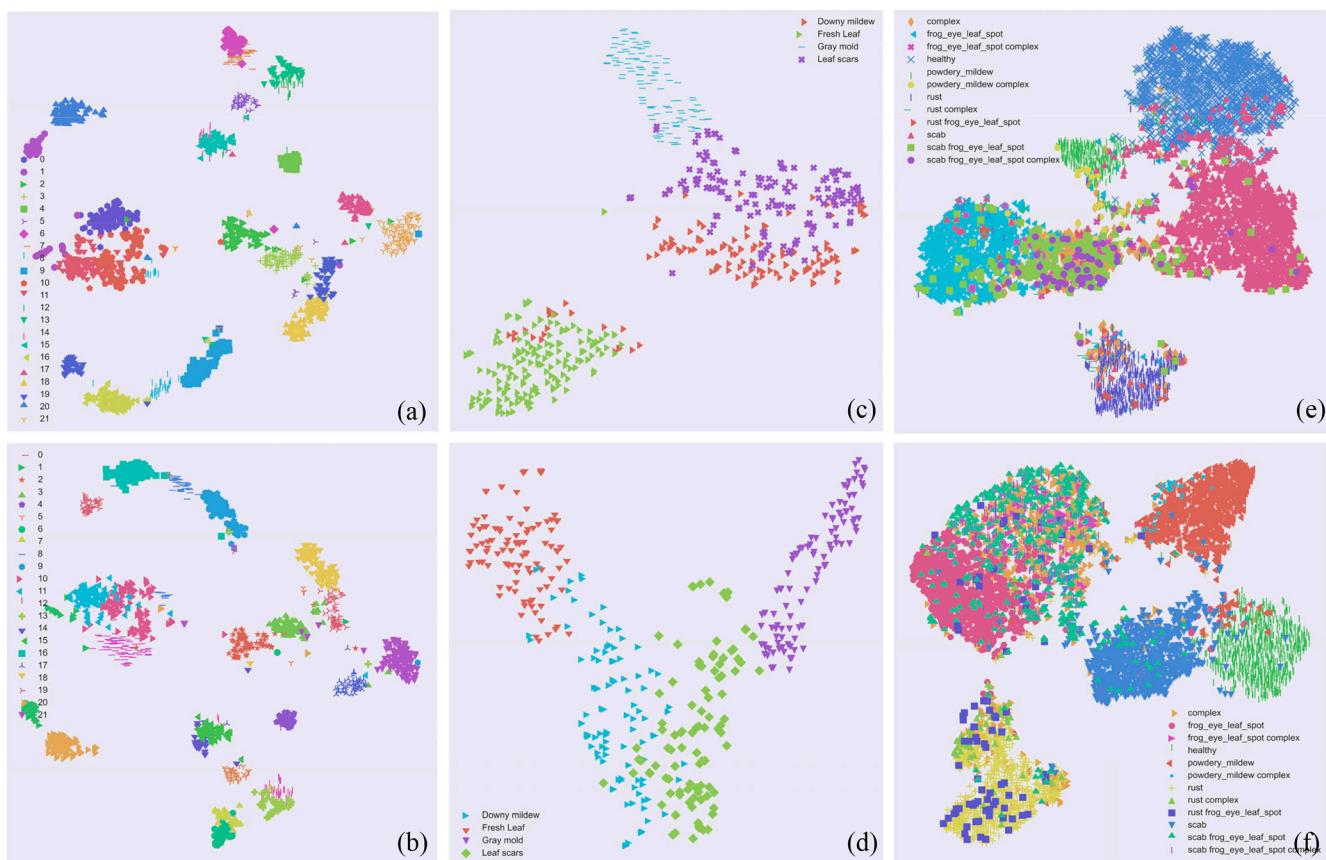


Fig. 7. Two-dimensional feature visualization plots of DFN-PSAN and ordinary classification DFN models in three data sets. (a) Feature visualization of DFN model in KatraTwelve, (b) Feature visualization of DFN-PSAN model in KatraTwelve, (c) Feature visualization of DFN model in BARI-Sunflower, (d) Feature visualization of DFN-PSAN model in BARI-Sunflower, (e) Feature visualization of DFN model in FGVC8, (f) Feature visualization of DFN-PSAN model in FGVC8.

clusters, the sample points in Fig. 7 (f) are more adjacent. This distinction between individual clusters will be more evident if the corresponding data are mapped to a high-dimensional space.

3.3. Comparative analysis of DFN structures

The Backbone, the backbone network of DFN, provides a rich semantic source for subsequent feature combination and aggregation. Existing studies have proposed improving Backbone to improve model performance, adapting to particular scenarios by replacing other feature extraction networks, or customizing one of the Backbone modules to improve recognition accuracy and lightweight (Chen et al., 2023). Li et al. (2023) proposed incorporating PicoDet into Backbone to streamline the model's parameters. Xu et al. (2023) tried to improve the

Backbone of YOLOv5 using a visual Transformer to obtain high accuracy. Similarly, the SwinTransformer encoder module was integrated into the YOLOv5 Backbone to achieve the highest accuracy (Lu et al., 2022). In order to compare the performance of DFN networks of the DFN-PSAN model in choosing different Backbone, we choose the popular network structure as much as possible. The Backbone networks that can be borrowed can be selected from three aspects. Classical neural networks choose EfficientNetV2, MobileNetV3, ShuffleNetV2, ResNet50, GhostConv. more novel neural networks choose RegNet, ConvNeXt, and HorNet. For networks based on the Transformer idea, choose SwinTransformer, CoTNet, and C3Transformer. It is worth noting that the Neck structure is consistent when DFN replaces the Backbone network, and all models choose the network structure with a smaller number of parameters as much as possible to ensure the

Table 7
Comparative test results of DFN-PSAN replacing other Backbone networks.

Model Type	Backbone Network	Accuracy (%) Katra-Twelve	BARI-Sunflower	FGVC8	Params	GFLOPs
Classic Network	EfficientNetV2	98.36	94.88	94.12	6,232,060	9.8
	MobileNetV3	96.89	93.76	85.81	3,151,520	3.1
	ShuffleNetV2	98.41	96.32	93.94	3,395,541	6.6
	ResNet50	97.84	95.46	94.64	11,909,893	15.63
	GhostConv	93.48	91.26	89.48	1,569,216	3.9
Novelty Network	RegNet	98.57	95.36	92.63	5,105,725	9.4
	ConvNeXt	99.26	96.11	94.87	112,651,485	76.2
	HorNet	99.74	95.25	95.33	88,140,416	66.1
Transformer Network	SwinTransformer	95.48	82.59	97.36	1,857,709	23.5
	CoTNet	97.24	75.77	93.49	1,754,208	4.2
	C3Transformer	89.59	78.54	90.55	1,748,176	3.8
	Our	98.37	94.23	93.24	1,757,152	4.2

comparability of the experiments.

The experimental results of DFN-PSAN replacing Backbone are shown in Table 7. On the Katra-Twelve dataset, the novel network type performs the best, with an average accuracy above 99.2 %. The combined model complexity comparison shows that the Transformer family of networks performs the worst, with SwinTransformer having a 2.36 % lower accuracy than ResNet50, while GFLOPs improve by 7.87. Some classical network types perform in line with the proposed DFN-PSAN performance but have more parameters with GFLOPs. For the BARI-Sunflower and FGVC8 datasets, the three types of models perform in line with the BARI-Sunflower and Katra-Twelve performance. However, the latter's accuracy is more robust than the former. This is due to the plain disease image background of the latter. The backbone of the Transformer series has the best performance on the FGVC8 dataset and the worst performance on the BARI-Sunflower dataset; the accuracy difference between SwinTransformer on the FGVC8 and BARI-Sunflower datasets is 14.77 %. Comparing the classical network type GhostConv and the novel network type RegNet, which have the worst performance among the same type of networks, the accuracy gap between them on the FGVC8 and BARI-Sunflower datasets is 1.78 % and 2.73 %, respectively, but the error rate is at least 12.04 % lower compared to that of SwinTransformer. Through experiments, it was found that the Transformer series of Backbone performed best on datasets with a large number of samples, FGVC8 was more complex compared to the other two datasets, but the network with Transformer coding module proved to overcome the ambiguous patterns not extracted by the convolutional neural network. At the same time, its higher computational cost consumption with a large number of the requirement of dataset construction is also a significant challenge for real agricultural plant disease identification scenarios. Collectively, our proposed DFN-PSAN model is leading on average.

3.4. Analysis of attentional mechanisms

The attention mechanism originated from the study of human vision, which can focus on the focused target region, suppress unimportant information, and improve the performance of deep learning models. The attention mechanism has been widely used in deep convolutional neural network models. We incorporated the PSA attention mechanism in the PSAN network architecture of the DFN-PSAN model. To explore the impact of other attention mechanisms on the model's performance, the attention mechanism structure that can be flexibly integrated into the PSAN network architecture and is widely used in academia was selected for testing. Specifically, the Squeeze-and-Excitation (SE) module, which adds an attention mechanism in the channel dimension, the Effective Squeeze-Excitation (ESE) module, which solves the information loss problem in the process of dimension increase and decrease, the Efficient Efficient Channel Attention (ECA) module, which performs adaptive attention modeling for each channel, and Convolutional block attention module (CBAM), which combines channel attention and spatial attention through a tandem structure. The ParNet is a novel attention module (Goyal et al., 2022), and this parallel structure allows the model to capture multi-scale feature information better, improving its representational power and performance. The lightweight channel attention mechanism Coordinate Attention (CA) is a method that can consider inter-channel relationships and location information over long distances. It effectively enhances the model's ability to perceive channel information while reducing computational consumption.

Based on these seven attention modules, seven sets of experiments were designed to compare the performance and accuracy of the model after adding attention, and the DFN of the DFN-PSAN model needed to be kept consistent in the comparison experiments to ensure the consistency of the results. The experimental data set is FGVC8, with the most complex disease spot characteristics and background, and the above seven attention structures are embedded into the PSAN network structure of the DFN-PSAN model for testing. Fig. 8 shows the experimental

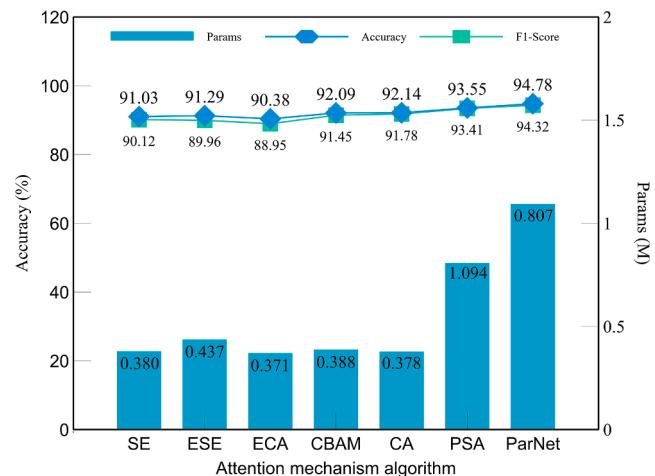


Fig. 8. Classification performance of the DFN-PSAN model applying seven attention mechanisms.

results. The ParNet attention mechanism achieved the highest accuracy of 94.78 %, and the ECA attention mechanism achieved the lowest accuracy of 90.38 %, with a difference of 4.4 % between the highest and lowest accuracies. The PSA attention mechanism used by DFN-PSAN achieved a suboptimal accuracy of 93.55 %, with a 1.23 % loss in accuracy compared to ParNet, but a 26 % savings in the number of model parameters. The CA and CBAM perform better than the spatial and channel-based attention mechanisms SE, ESE, and ECA because they are advantageous for plant disease identification at a fine-grained scale by stacking modules and using contextual information to improve feature representation gradually. The PSA highlights important regions in plant disease images through a pixel-level attention mechanism is the key to achieving better recognition accuracy.

3.5. Comparative analysis of the current state of technology

In order to compare the proposed DFN-PSAN, we discuss in Table 8 some recent work that is particularly useful to explore the advantages of other different model architectures. The experimental data in this paper were tested using three relatively new datasets closely related to real agricultural scenarios, and the results obtained, although not optimal, reflect a competitive performance compared to other state-of-the-art models. In Table 8, the performance of DFN-PSAN in the Katra-Twelve dataset is lower than that of the PPLC-Net model. However, the accuracy is improved by at least 0.56 % compared to the other three models. In contrast, for BARI-Sunflower and FGVC8, the performance is improved to varying degrees compared to the other advanced methods. Across the board the performance of the DFN-PSAN model in BARI-Sunflower and FGVC8 was weaker than that of Katra-Twelve, which was mainly attributed to the different shooting angles and rich shooting backgrounds of the former, which were more potent than the latter in accurately reflecting the complex disease symptoms of plants. Regarding studies based on other datasets, most studies chose the open-source database PlantVillage; a few studies differed part of the data from PlantVillage by data augmentation or added homemade data to form a new dataset, and very few studies validated by homemade datasets. To form a comparison with other research in the field of plant disease identification, our proposed DFN-PSAN model obtained 99.89 % accuracy in PlantVillage, which is slightly lower than the 0.05 % accuracy of the classical neural network model Inception embedded in Visual Transformer (ICVT). However, it reflects the accuracy as a general Convolutional Neural Network the DFN-PSAN advantage. In conclusion, in both the image feature fusion multi-headed network and the neural network incorporating attention mechanism, DFN plays a crucial role in its feature extraction automation method composed of Backbone and

Table 8

Results of recent baseline studies and other relevant studies on three datasets.

Dataset	Research	Model	No. of classes	Performance (%)
Katra-Twelve	Tiwari, Joshi, and Dutta (2022)	DenseNet201	22	97.69
	Russel and Selvaraj (2022)	Laws mask + Multi-streaming CNN		97.16
	Gehlot and Gandhi (2023)	EffiNet-TS		95.74
	Dai, Fan et al. (2023)	PPLC-Net		98.44
	Ours	DFN-PSAN		98.25
BARI-Sunflower	Ghosh, Mondal et al. (2023)	VGG19 + CNN	4	93.00
	Ours	DFN-PSAN		94.47
FGVC8	Yadav et al. (2022)	AFD-Net	6	92.60
	Ours	DFN-PSAN		93.55
Others	Gokulnath and Usha Devi (2021)	LF-CNN	5	98.93
	Chen, Zhang et al. (2021)	MobileNet-V2 + Attention	38	99.67
	Sunil et al. (2022)	EfficientNetV2	3	98.26
	Chen, Duan et al. (2022)	MS-DNet	38	98.32
	Sanida et al. (2022)	DenseNet201	38	99.87
	Nandhini and Ashokkumar (2022)	M-HGSO + Densenet-121	14	98.70
	Yu,Xie, and Huang (2023)	ICVT(Transformer)	38	99.94
	Kaya and Gürsoy (2023)	DenseNet121 + Imagefusion	38	98.12
	Diana Andrusia et al. (2023)	Capsule-network	5	99.12
	Ours	DFN-PSAN(PlantVillage)	38	99.89

Neck.

3.6. SHAP interpretability analysis

In this section, the DFN-PSAN model will be analyzed for its power of feature attention in identifying different plant disease images with different disease categories when faced with correctly identified samples versus incorrectly identified samples. Due to the poor interpretability of most deep learning algorithms, it is not possible to analyze how individual features affect the prediction results of the model, which is very

important for plant disease diagnosis. Therefore, this paper uses the SHAP method to analyze the factors influencing the different disease category features in the model. The Shapley Additive exPlanation (SHAP) is an additive explanatory model based on game theory. Its basic idea is to explain the model by calculating the marginal contribution of each feature when it is added to the model to explain the model, which can be used to explain various black box models (Ji et al., 2022). The method first calculates the contribution value of each feature, which may be positive or negative, and then accumulates the contribution values of all features to obtain the final prediction.

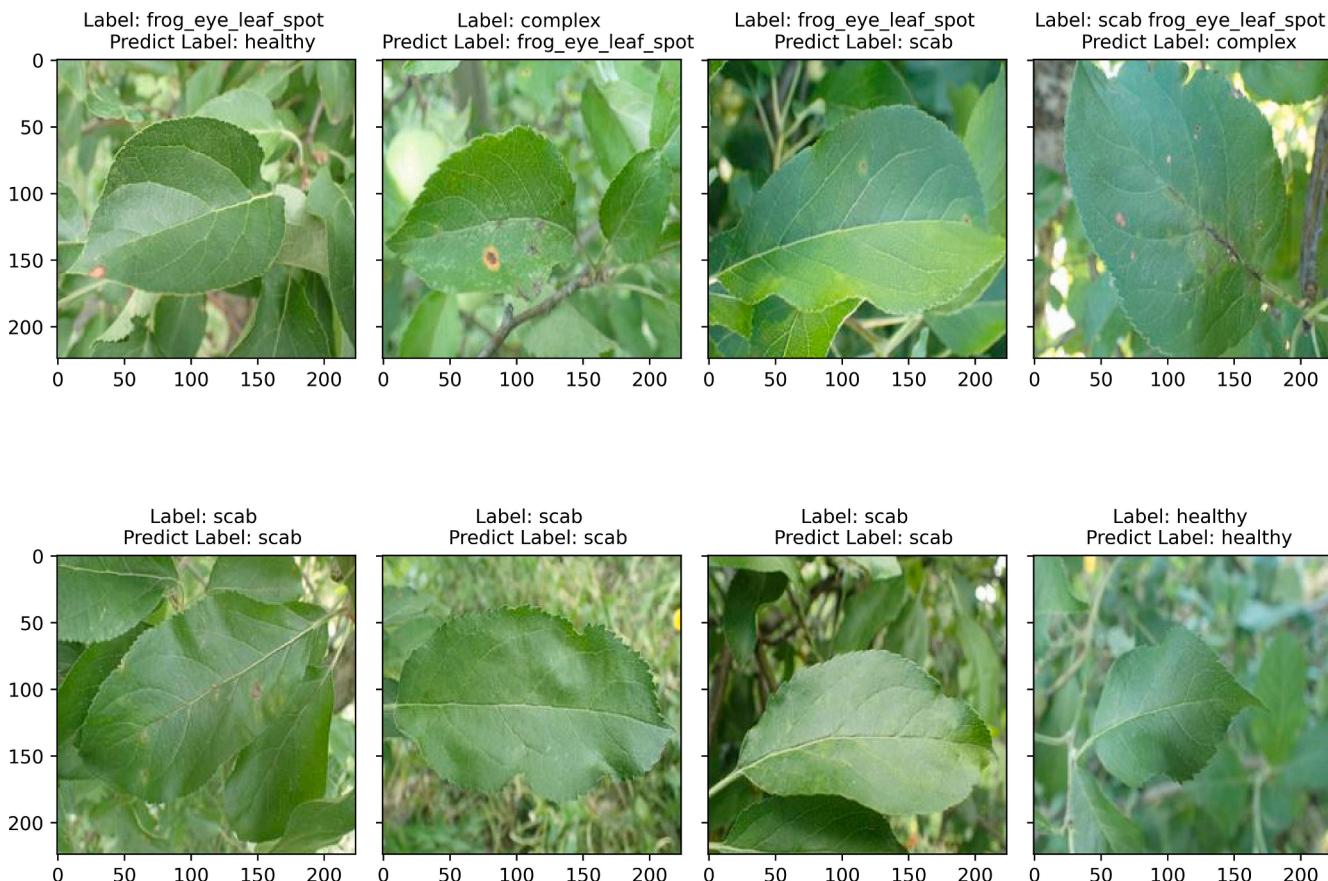


Fig. 9. True labels and prediction results of the DFN-PSAN model for random test samples in the test set. (a) samples with incorrect predictions, (b) samples with correct predictions.

As shown in Fig. 9, to implement this process, the FGVC8 dataset with the most complex disease pattern collected from the natural field environment is selected, followed by randomly selected images from the test set of FGVC8 for recognition, and the actual labels and predicted results are displayed above the obtained result images, respectively. From the result images, two images with correct identification and incorrect identification are selected for SHAP calculation, respectively, and the results are converted into heat maps to realize two-dimensional visualization and analysis. Fig. 10 and Fig. 11 show the incorrectly identified and correctly identified SHAP images, respectively. The blue gradient block indicates negative contribution, the red gradient block indicates positive contribution, and the blue box indicates positive and negative cumulative contribution influence.

The characteristic attention concerns of the DFN-PSAN model for misidentifying plant disease categories are shown in Fig. 10. The sample shown in Fig. 10 (a) is correctly labeled with a disease category that should be FLS but is incorrectly identified by the model as a healthy category based on the analysis in Fig. 9 (a). Comparing the heat map of FLS with the health category in detail, the blue box incorrectly identified by the model as the health category is significantly larger than the real category FLS blue box, which is why the model incorrectly considers the health category. It can be observed that the healthy category has two

more regions of blue gradient blocks compared to FLS. Since the blue gradient blocks point to a negative contribution, it proves that the features of these two regions do not determine the model identification of FLS diseases. In addition, the blue-red gradient blocks possessed by FLS have distributional similarity with those of the other nine disease categories, indicating that there are more cross-feature points in different disease categories, which makes model identification more difficult. For Fig. 10 (b), the correct labeled disease categories should be scab and FLS. However, the model incorrectly identifies them as complex disease categories based on the analysis of Fig. 9 (b). The correct category in Fig. 10 (b) has distributional similarity with the red asymptotic blocks of the other six disease categories. It should be noted that they all contribute positively. The difference in SHAP positive contribution values is slight, which intuitively reflects the difficulty of DFN-PSAN in identifying complex features.

The characteristic attention concerns of the DFN-PSAN model to correctly identify plant disease categories are shown in Fig. 11. The correct labeled disease category for the samples shown in Fig. 11 (a) and Fig. 11 (b) should be the scab, and based on the analysis of Fig. 9 (a), they both correctly identified the scab category. The blue boxes in Fig. 11 (a) and Fig. 11 (b) are more significant compared to Fig. 10, indicating that the contribution influence allows the model to have a

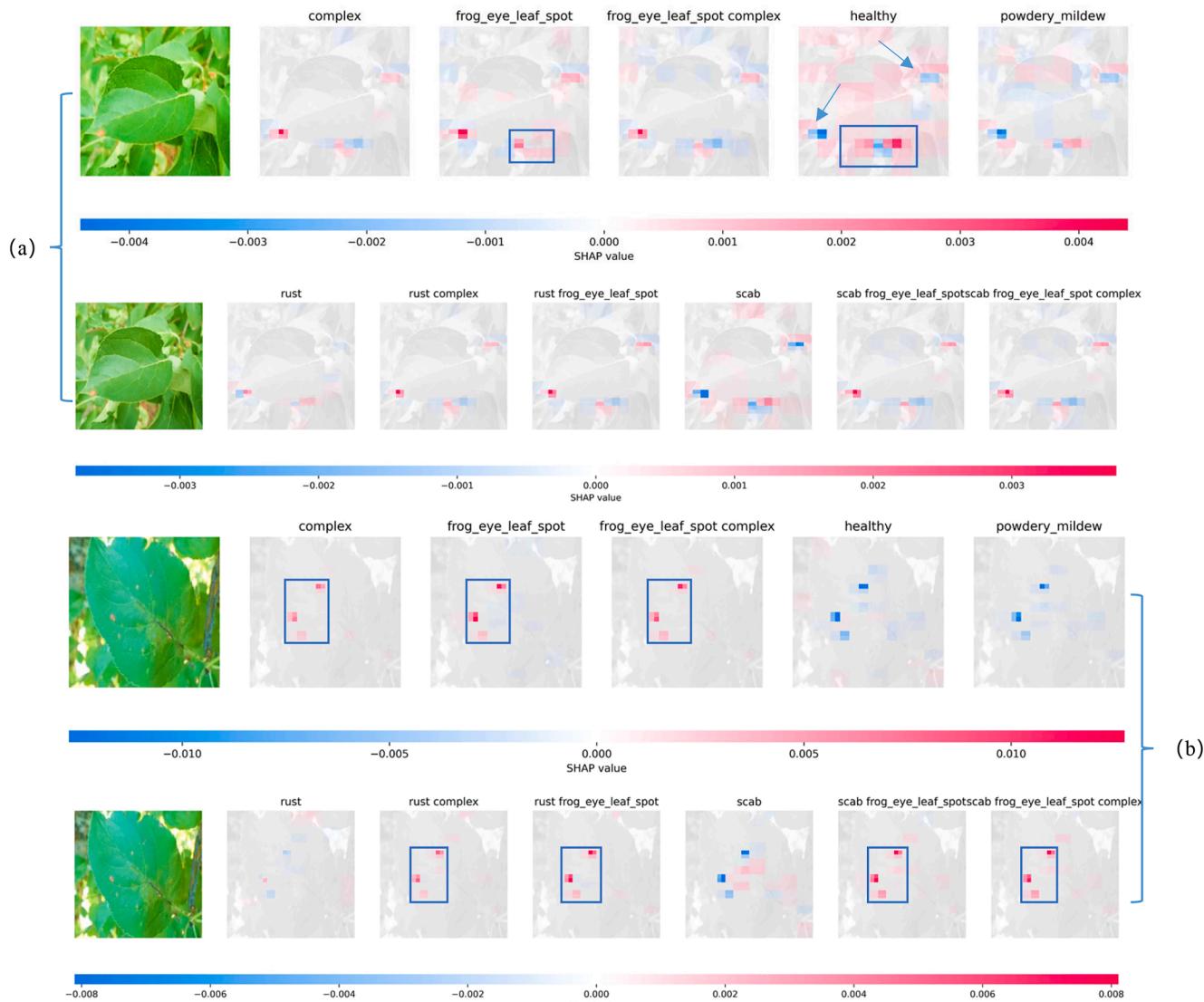


Fig. 10. Explaining the attentional feature concerns of the DFN-PSAN model for misidentified plant diseases. (a) the analysis results of the first image on the left side of Fig. 9(a), (b) the analysis results of the first image on the right side of Fig. 9(a).

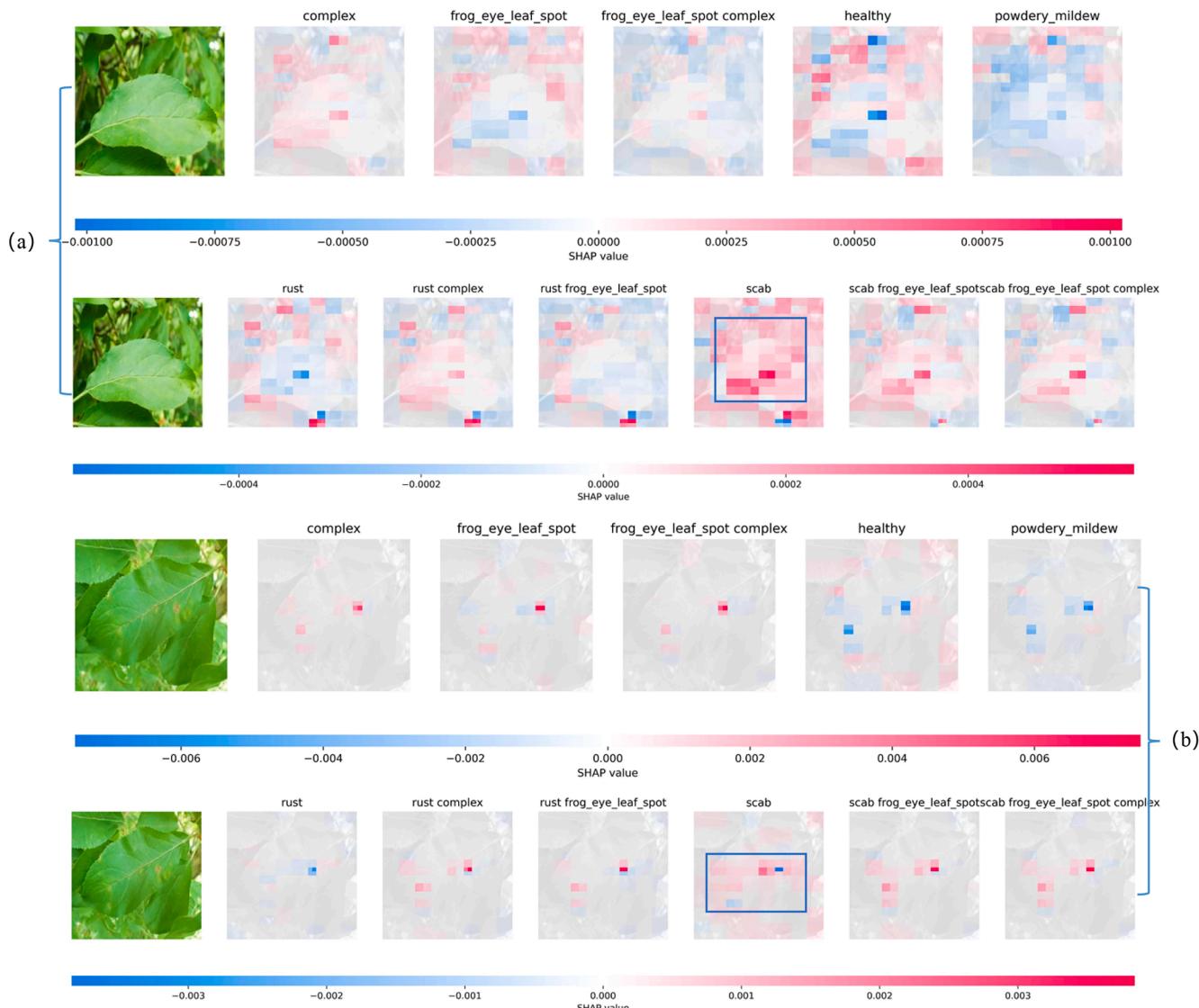


Fig. 11. Explaining the attentional feature concerns of the DFN-PSAN model for correct identification of plant diseases. (a) The analysis results of the third image on the left side of Fig. 9(b), (b) the analysis results of the first image on the left side of Fig. 9(b).

higher confidence value in determining the disease category, thus allowing the model to identify the disease type. Fig. 11 (a) The red blocks of the heat map of the scab disease category cover almost the whole leaf area. In contrast, the blue blocks cover all the disease categories except the complex category disease, indicating that DFN-PSAN is more sensitive to the characteristics of the scab disease category. Also, combining the image analysis results of both categories in Fig. 11 (b) and Fig. 10 shows that the complex disease category has some interference effect on the model to discriminate other disease categories. Also, it indicates that the plant pathologists define more features of other disease categories or directly include other disease categories for the complex disease category when labeling the complex disease category.

4. Discussion

The YOLOv5 family of neural network models is widely used as the basis for most research, as it achieves a balance between detection speed and accuracy and supports a variety of optimal deployment methods. Initially popular in areas related to computer vision target detection, it has gained initial support in classification, semantic, and instance segmentation applications. However, the studies using the YOLOv5 series network as the target method in plant disease classification studies are

less involved. In the current study, improving YOLOv5 series networks mainly focuses on improving and optimizing the Backbone and Neck. For example, the novel ConvNeXt, SwinTransformer, and CoTNet networks are used to improve Backbone, which can enhance the extraction of features as a core component. Improved attention mechanisms CBAM, SE, and ECA modules are usually located at the tail of the Backbone for enhanced extraction of disease feature details (Lu et al., 2022; Qi et al., 2022). For Neck, which has a feature pyramid network (FPN) and a path aggregation network (PAN), the available choice is to use a weighted bi-directional feature pyramid network (BiFPN) or to incorporate an attention mechanism for optimization (Li et al., 2022; Xu et al., 2023). These targeted improvements have achieved better results, but to research the classification algorithm of YOLOv5, the proposed DFN-PSAN model retains the YOLOv5 infrastructure and chooses the most parsimonious approach to improve Head. Compared with DFN-PSAN, Yang, Bist et al. (2023) used EfficientNetV2 to improve Backbone and achieve chicken behavior classification with only Head, i.e., the original classification network structure. Since the original classification network structure is too simple, extracting useful information in overly complex feature patterns is not easy. Therefore, the proposed PSAN retains the Neck structure in the target detection network, is designed to replace the Head structure, and chooses the lightweight Backbone with

the Neck as the source of subsequent classification feature patterns, balancing the requirements of computational resource consumption and higher accuracy.

The current need to move closer to real agricultural scenarios in plant disease research has led to increasingly complex disease identification and higher demands on model performance. The proposed DFN-PSAN model replacing other Backbones achieved competitive performance in comparison tests. However, the comparative tests of the three datasets in Table 7 illustrate that the accuracy of classical neural networks in real complex agricultural scenarios is not necessarily worse than the newer neural networks based on the Transformer idea. The model selection needs to consider the relationship between fuzzy feature patterns and sample capacity. Besides, the recognition accuracy of DFN-PSAN in three plant disease datasets by data augmentation technique was improved to different degrees (Table 6), and the more complex the fine-grained classification feature pattern is, the better the performance is after the second-stage data augmentation. Combined with the analysis in Table 7, the visual Transformer family of neural networks has a high dependence on the sample size of the dataset but has a significant advantage in extracting complex disease feature patterns. When facing research scenarios with a large number of datasets, the optimized Transformer neural network usually has better performance in the context of plain data, and the proposed DFN-PSAN, although with better overall performance, adopts the introduction of the feature aggregation network, the attention mechanism and the data augmentation method to assist in improving the performance, and thus faces a particular disadvantage in terms of the generality of the model. Overall, leaf disease prediction in smart agriculture has significant potential benefits. However, the disease data analysis and model design become critical factors influenced by several factors such as computational speed, treatment cost, and recognition accuracy.

5. Conclusion

This work presents a novel lightweight deep-learning model to identify diseases in plant leaves. Explicitly, we propose a DL framework called the DFN-PSAN approach, which preserves the Backbone and Neck structure of YOLOv5 as the core component of the feature extraction network, designs a neural network with an attention mechanism to reconstruct the YOLOv5 classification layer, and incorporates the PSA attention mechanism to highlight important regions in plant disease images from the pixel level. Thus, a representative set of feature patterns can be extracted. Next, the two-stage weather data augmentation method improves the generalization ability of three real agricultural scenarios plant disease datasets in the DFN-PSAN model, and the t-SNE and SHAP methods explain the feature influencing factors of the model in terms of both feature layer data dimension and model attention. The results showed that the two-stage weather data augmentation improved the accuracy of the three plant disease datasets by at least 0.69 %, and the PSA attention mechanism employed by DFN-PSAN achieved a sub-optimal accuracy of 93.55 %, with a 26 % savings in the number of model parameters compared to the optimal attention mechanism. In conclusion, the average accuracy of the proposed DFN-PSAN model exceeded 95.27 % in all three datasets, proving that the model can effectively classify a wide range of plant diseases in real agricultural scenarios.

Future research will involve building pre-trained neural network model weights for extensive data of plant pests and diseases for real agricultural scenarios to help other models converge faster when replacing the feature extraction network Backbone. This work could further alleviate the difficulties of pest and disease identification in intelligent agriculture.

CRediT authorship contribution statement

Guowei Dai: Conceptualization, Methodology, Validation,

Investigation, Data curation, Writing – original draft, Writing – review & editing, Formal analysis, Project administration. **Zhimin Tian:** Writing – review & editing. **Jingchao Fan:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **C.K. Sunil:** Writing – review & editing. **Christine Dewi:** Writing – review & editing.

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.

Acknowledgments

This work was supported by the National Agriculture Science Data Center (NASDC2023XM00). The authors thank the funding agency for financial support. The authors would also like to thank all authors and anonymous reviewers cited in this paper for their helpful comments and suggestions.

References

- Alirezazadeh, P., Schirrmann, M., Stolzenburg, F., 2023. Improving deep learning-based plant disease classification with attention mechanism. *Gesunde Pflanzen* 75 (1), 49–59. <https://doi.org/10.1007/s10343-022-00796-y>.
- B.v., G., & G., U. D. (2021). Identifying and classifying plant disease using resilient LF-CNN. *Ecological Informatics*, 63, 101283. <https://doi.org/10.1016/j.ecoinf.2021.101283>.
- C. K., S., C. D., J., & Patil, N. (2022). Cardamom Plant Disease Detection Approach Using EfficientNetV2. *IEEE Access*, 10, 789–804. <https://doi.org/10.1109/ACCESS.2022.13138920>.
- C.k., S., C.d., J., & Patil, N. (2023). Tomato plant disease classification using Multilevel Feature Fusion with adaptive channel spatial and pixel attention mechanism. *Expert Systems with Applications*, 228, 120381. <https://doi.org/10.1016/j.eswa.2023.120381>.
- Chen, W., Chen, J., Duan, R., Fang, Y., Ruan, Q., Zhang, D., 2022. MS-DNet: A mobile neural network for plant disease identification. *Comput. Electron. Agric.* 199, 107175 <https://doi.org/10.1016/j.compag.2022.107175>.
- Chen, S., Liao, Y., Lin, F., Huang, B., 2023. An improved lightweight YOLOv5 algorithm for detecting strawberry diseases. *IEEE Access* 11, 54080–54092. <https://doi.org/10.1109/ACCESS.2023.3282309>.
- Chen, J., Zhang, D., Zeb, A., Nanehkaran, Y.A., 2021. Identification of rice plant diseases using lightweight attention networks. *Expert Syst. Appl.* 169, 114514 <https://doi.org/10.1016/j.eswa.2020.114514>.
- Chouhan, S.S., Singh, U.P., Kaul, A., Jain, S., 2019. A data repository of leaf images: Practice towards plant conservation with plant pathology. In: 2019 4th International Conference on Information Systems and Computer Networks (ISCON), pp. 700–707. <https://doi.org/10.1109/ISCON47742.2019.9036158>.
- Dai, G., Fan, J., Tian, Z., Wang, C., 2023. PLLC-Net: Neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism. *Journal of King Saud University - Computer and Information Sciences* 35 (5), 101555. <https://doi.org/10.1016/j.jksuci.2023.101555>.
- Diana Andrushia, A., Mary Neelha, T., Trepheena Patricia, A., Umadevi, S., Anand, N., Varshney, A., 2023. Image-based disease classification in grape leaves using convolutional capsule network. *Soft. Comput.* 27 (3), 1457–1470. <https://doi.org/10.1007/s00500-022-07446-5>.
- Dogra, R., Rani, S., Singh, A., Albahar, M.A., Barrera, A.E., Alkhayyat, A., 2023. Deep learning model for detection of brown spot rice leaf disease with smart agriculture. *Comput. Electr. Eng.* 109, 108659 <https://doi.org/10.1016/j.compeleceng.2023.108659>.
- Fang, X., Zhen, T., Li, Z., 2023. Lightweight multiscale CNN model for wheat disease detection. *Appl. Sci.* 13 (9), Article 9. <https://doi.org/10.3390/app13095801>.
- Gao, F., Luo, X., Yang, Z., Zhang, Q., 2022. Label smoothing and task-adaptive loss function based on prototype network for few-shot learning. *Neural Netw.* 156, 39–48. <https://doi.org/10.1016/j.neunet.2022.09.018>.
- Gehlot, M., Gandhi, G.C., 2023. “EffiNet-TS”: A deep interpretable architecture using EfficientNet for plant disease detection and visualization. *J. Plant Dis. Prot.* 130 (2), 413–430. <https://doi.org/10.1007/s41348-023-00707-x>.
- Ghosh, P., Mondal, A.K., Chatterjee, S., Masud, M., Meshref, H., Bairagi, A.K., 2023. Recognition of sunflower diseases using hybrid deep learning and its explainability with AI. *Mathematics* 11 (10), Article 10. <https://doi.org/10.3390/math11102241>.
- Goyal, A., Bochkovskiy, A., Deng, J., & Koltun, V. (2022). Non-deep Networks. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in Neural Information Processing Systems* (Vol. 35, pp. 6789–6801). Curran Associates,

- Inc. https://proceedings.neurips.cc/paper_files/paper/2022/file/2d52879ef2ba487445ca2e143b104c3b-Paper-Conference.pdf.
- Hassan, S.M., Maji, A.K., 2022. Plant Disease Identification Using a Novel Convolutional Neural Network. IEEE Access 10, 5390–5401. <https://doi.org/10.1109/ACCESS.2022.3141371>.
- Jayagopal, P., Rajendran, S., Mathivanan, S.K., Sathish Kumar, S.D., Raja, K.T., Paneerselvam, S., 2022. Identifying region specific seasonal crop for leaf borne diseases by utilizing deep learning techniques. Acta Geophys. 70 (6), 2841–2854. <https://doi.org/10.1007/s11600-022-00759-x>.
- Ji, S., Wang, X., Lyu, T., Liu, X., Wang, Y., Heinen, E., Sun, Z., 2022. Understanding cycling distance according to the prediction of the XGBoost and the interpretation of SHAP: A non-linear and interaction effect analysis. J. Transp. Geogr. 103, 103414 <https://doi.org/10.1016/j.jtrangeo.2022.103414>.
- Kaya, Y., Gürsoy, E., 2023. A novel multi-head CNN design to identify plant diseases using the fusion of RGB images. Eco. Inform. 75, 101998 <https://doi.org/10.1016/j.ecoinf.2023.101998>.
- Khan, A.I., Quadri, S.M.K., Banday, S., Latief Shah, J., 2022. Deep diagnosis: A real-time apple leaf disease detection system based on deep learning. Comput. Electron. Agric. 198, 107093 <https://doi.org/10.1016/j.compag.2022.107093>.
- Legrand, N., 2023. War in Ukraine: The rational “wait-and-see” mode of global food markets. Appl. Econ. Perspect. Policy 45 (2), 626–644. <https://doi.org/10.1002/aep.13335>.
- Li, K., Li, X., Liu, B., Ge, C., Zhang, Y., Chen, L., 2023. Diagnosis and application of rice diseases based on deep learning. PeerJ Comput. Sci. 9, e1384.
- Li, J., Qiao, Y., Liu, S., Zhang, J., Yang, Z., Wang, M., 2022. An improved YOLOv5-based vegetable disease detection method. Comput. Electron. Agric. 202, 107345 <https://doi.org/10.1016/j.compag.2022.107345>.
- Lu, S., Liu, X., He, Z., Zhang, X., Liu, W., Karkee, M., 2022. Swin-Transformer-YOLOv5 for Real-Time Wine Grape Bunch Detection. Remote Sens. (Basel) 14 (22), Article 22. <https://doi.org/10.3390/rs14225853>.
- Nandhini, S., Ashokkumar, K., 2022. An automatic plant leaf disease identification using DenseNet-121 architecture with a mutation-based Henry gas solubility optimization algorithm. Neural Comput. & Applic. 34 (7), 5513–5534. <https://doi.org/10.1007/s00521-021-06714-z>.
- Ni, J., Zhou, Z., Zhao, Y., Han, Z., & Zhao, L. (2023). Tomato leaf disease recognition based on improved convolutional neural network with attention mechanism. *Plant Pathology, n/a(n/a)*. <https://doi.org/10.1111/ppa.13745>.
- Pavithra, A., Kalpana, G., Vigneswaran, T., 2023. Deep learning-based automated disease detection and classification model for precision agriculture. Soft. Comput. <https://doi.org/10.1007/s00500-023-07936-0>.
- Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., Li, M., Bao, Z., Li, Y., 2022. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. Comput. Electron. Agric. 194, 106780 <https://doi.org/10.1016/j.compag.2022.106780>.
- Reddy, S.R.G., Varma, G.P.S., Davuluri, R.L., 2023. Resnet-based modified red deer optimization with DCLNN classifier for plant disease identification and classification. Comput. Electr. Eng. 105, 108492 <https://doi.org/10.1016/j.compeleceng.2022.108492>.
- Russel, N.S., Selvaraj, A., 2022. Leaf species and disease classification using multiscale parallel deep CNN architecture. Neural Comput. & Applic. 34 (21), 19217–19237. <https://doi.org/10.1007/s00521-022-07521-w>.
- Sanida, T., Tsiktsiris, D., Sideris, A., Dasgupta, M., 2022. A heterogeneous implementation for plant disease identification using deep learning. Multimed. Tools Appl. 81 (11), 15041–15059. <https://doi.org/10.1007/s11042-022-12461-7>.
- Sara, U., Rajbongshi, A., Shakil, R., Akter, B., Sazzad, S., Uddin, M.S., 2022. An extensive sunflower dataset representation for successful identification and classification of sunflower diseases. Data Brief 42, 108043. <https://doi.org/10.1016/j.dib.2022.108043>.
- Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A., 2022. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Syst. Appl. 208, 118117 <https://doi.org/10.1016/j.eswa.2022.118117>.
- Tiwari, V., Joshi, R.C., Dutta, M.K., 2022. Deep neural network for multi-class classification of medicinal plant leaves. Expert. Syst. 39 (8), e13041.
- Wani, J.A., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S., Singh, S., 2022. Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: methodologies, applications, and challenges. Arch. Comput. Meth. Eng. 29 (1), 641–677. <https://doi.org/10.1007/s11831-021-09588-5>.
- Xu, J., Yao, J., Zhai, H., Li, Q., Xu, Q., Xiang, Y., Liu, Y., Liu, T., Ma, H., Mao, Y., Wu, F., Wang, Q., Feng, X., Mu, J., Lu, Y., 2023. TrichomeYOLO: A neural network for automatic maize trichome counting. Plant Phenomics 5, 0024. <https://doi.org/10.34133/plantphenomics.0024>.
- Yadav, A., Thakur, U., Saxena, R., Pal, V., Bhatia, V., Lin, J.-C.-W., 2022. AFD-Net: Apple Foliar Disease multi classification using deep learning on plant pathology dataset. Plant and Soil 477 (1), 595–611. <https://doi.org/10.1007/s11104-022-05407-3>.
- Yang, X., Bist, R., Subedi, S., Wu, Z., Liu, T., Chai, L., 2023. An automatic classifier for monitoring applied behaviors of cage-free laying hens with deep learning. Eng. Appl. Artif. Intel. 123, 106377 <https://doi.org/10.1016/j.engappai.2023.106377>.
- Yu, M., Ma, X., Guan, H., 2023a. Recognition method of soybean leaf diseases using residual neural network based on transfer learning. Eco. Inform. 76, 102096 <https://doi.org/10.1016/j.ecoinf.2023.102096>.
- Yu, S., Xie, L., Huang, Q., 2023b. Inception convolutional vision transformers for plant disease identification. Internet of Things 21, 100650. <https://doi.org/10.1016/j.iot.2022.100650>.
- Zhang, H., Zu, K., Lu, J., Zou, Y., & Meng, D. (2022). EPSANet: An Efficient Pyramid Squeeze Attention Block on Convolutional Neural Network. 1161–1177. <https://doi.org/10.48550/arXiv.2105.14447>.
- Zhu, X., Li, J., Jia, R., Liu, B., Yao, Z., Yuan, A., Huo, Y., Zhang, H., 2023. LAD-Net: A Novel Light Weight Model for Early Apple Leaf Pests and Diseases Classification. IEEE/ACM Trans. Comput. Biol. Bioinf. 20 (2), 1156–1169. <https://doi.org/10.1109/TCBB.2022.3191854>.



GUOWEI DAI received his B.S. degree in Computer Science and Technology from Wuhan Polytechnic University in 2021, and is currently pursuing his M.S. degree at the Agricultural Information Institute, Chinese Academy of Agricultural Sciences. His main research interests are agricultural artificial intelligence and agricultural information technology. He published more than 8 research papers, among which, 4 were indexed by SCI; he was granted 7 national invention patents and 3 software copyrights. He serves as a reviewer for Computers and Electronics in Agriculture, Precision Agriculture, Journal of Agricultural Engineering, The Journal of Engineering and Complex & Intelligent Systems. His contact email: dgwstyle@foxmail.com, ORCID: <https://orcid.org/0000-0003-0054-0392>.



ZHIMIN TIAN received the B.S. degree in Information and Computational Science from Hebei University of Engineering in 2011 and the M.S. degree in Computer Application Technology from Hebei University of Engineering in 2014. He is a senior laboratory technician in the School of Hebei University of Water Resources and Electric Engineering. He is mainly engaged in research on big data analysis and neural network technology. He has presided over and participated in 10 projects, including the Municipal department-level project. He has published six papers authorized four utility model patents and two software copyrights.



JINGCHAO FAN received the B.S. degree in Computer Science and Technology from Liaoning University in 2003, the M.S. degree in Management Science and Engineering from the Chinese Academy of Agricultural Sciences in 2006, and the Ph.D. degree in Crop Information Science from the Chinese Academy of Agricultural Sciences in 2011. He is currently the Deputy Director of Scientific Data Research Department, Agricultural Information Institute of CAAS (AIICAAS). He is mainly engaged in research on agricultural information technology, Chinese web information retrieval, and fruit tree model construction. He has presided over and participated in 7 projects such as 863, transformation of scientific and technological achievements, and basic scientific research business fund, etc. He has participated in the construction of 6 national science and technology infrastructure platforms as a technical backbone, etc. He has obtained 3 scientific and technological achievements, published 12 papers; authorized 15 utility model patents, applied for 4 invention patents; 16 software copyrights.



SUNIL C. K. (Student Member, IEEE) received a Ph.D. from the National Institute of Technology Karnataka Surathkal, India; he is currently with BNM Institute of Technology Bengaluru as an Associate Professor. His research area includes machine learning and deep learning applications in agriculture and allied areas and robotics.



CHRISTINE DEWI received a B.S. Degree (S. Kom.) from the Informatics Engineering study program in 2010, and a Master of Computer Science (M.Cs.) from the Master of Information Systems study program in 2012, both from the Faculty of Information Technology, Satya Wacana Christian University, Salatiga, Indonesia. In 2021 she finished her Ph.D. and work as a postdoctoral researcher at the College of Informatics, Chaoyang University of Technology, Taiwan until 2022. She received an Excellent Dissertation Award from the Taiwanese Association for Consumer Electronics (TACE), Taiwan 2021. This award aims to honor a dissertation project funded by MOST Taiwan. The participants were all students from various universities in Taiwan. She is now an Associate Professor in the Department of Information Technology and serves as Head of the Associate Degree Informatics Engineering Study Program at Satya Wacana Christian University, Indonesia. In 2022, she got AUTHOR WITH THE HIGHEST SINTA SCORE AND RESEARCH PUBLICATIONS in 3 Years AWARD from Satya Wacana Christian University, Indonesia. She is active as a reviewer and guest editor in many journals. Her current research interests include Image Processing, Computer Vision, Object Detection and Recognition, Artificial Intelligence, and Machine Learning.