frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 06 November 2025
DOI 10.3389/fpls.2025.1676798

OPEN ACCESS

EDITED BY
Bimlesh Kumar,
Indian Institute of Technology Guwahati, India

REVIEWED BY
Nevien Adel Ismaeil,
Al-Azhar University, Egypt
Feiyong He,
Macao Polytechnic University, Macao SAR, China

*CORRESPONDENCE
Bingjing Jia
jiabj@ahstu.edu.cn

RECEIVED 31 July 2025
ACCEPTED 29 September 2025
PUBLISHED 06 November 2025

CITATION
Zeng J, Jia B, Song C, Ge H, Shi L and Kang B (2025) CDPNet: a deformable ProtoPNet for interpretable wheat leaf disease identification. *Front. Plant Sci.* 16:1676798. doi: 10.3389/fpls.2025.1676798

CDPNet: a deformable ProtoPNet for interpretable wheat leaf disease identification

Jinyu Zeng[1], Bingjing Jia[1]*, Chenguang Song[1], Hua Ge[1], Lei Shi[2] and Bo Kang[1]

[1]College of Information and Network Engineering, Anhui Science and Technology University, Bengbu, Anhui, China, [2]State Key Laboratory of Media Convergence and Communication, Communication University of China, Beijing, China

Introduction: Accurate identification of wheat leaf diseases is crucial for food security, but existing prototype-based computer vision models struggle with the scattered nature of lesions in field conditions and lack interpretability.
Methods: To address this, we propose the Contrastive Deformable Prototypical part Network (CDPNet). The idea of CDPNet is to identify key image regions that influence model decisions by computing similarity measures between convolutional feature maps and latent prototype feature representations. Moreover, to effectively separate the disease target area from its complex background noise and enhance the discriminability of disease features, CDPNet introduces the Cross Attention (CA) Mechanism. Additionally, to address the scarcity of wheat leaf disease image data, we employ the Barlow Twins self-supervised contrastive learning method to capture feature differences across samples. This approach enhances the model's sensitivity to inter-class distinctions and intra-class consistency, thereby improving its ability to differentiate between various diseases.
Results: Experimental results demonstrate that the proposed CDPNet achieves an average recognition accuracy of 95.83% on the wheat leaf disease dataset, exceeding the baseline model by 2.35%.
Discussion: Compared to other models, this approach delivers superior performance and provides clinically interpretable decision support for the identification of real-world wheat diseases in field settings.

KEYWORDS

identification of wheat leaf diseases, interpretability, CDPNet, Cross Attention, Barlow Twins

1 Introduction

Wheat is one of the three major global food crops, ranking among the highest in both production volume and cultivated area. Its widespread cultivation and stable yields serve as a critical safeguard for global food security. However, disease infestation throughout its growth stages remains the primary challenge limiting stable and high yields (Bao et al.,

2021a). Statistics show that leaf diseases, such as leaf blight, mildew, and rust, can lead to annual global wheat yield losses ranging from 10% to 30%. These diseases not only lead to direct yield reductions but also trigger secondary hazards, such as grain quality deterioration and mycotoxin contamination, causing substantial losses in agricultural production (Simón et al., 2021). Therefore, accurate identification of wheat diseases, particularly leaf diseases, is critical for implementing effective control measures and ensuring healthy growth to enhance yields (Nigam et al., 2023).

With advancements in modern technology, machine learning and deep learning techniques are increasingly being applied to crop pest and disease detection. These techniques have shown highly promising results in achieving precise identification of crop pests and diseases using computer vision technology (Deng et al., 2025). Traditional machine learning techniques, such as Support Vector Machines (SVM) (Rezvani and Wu, 2023), Random Forests (Gao et al., 2022), and Decision Trees (Alaniz et al., 2021), have been widely employed in wheat disease detection. These techniques employ various algorithms to extract different features from images, including color, texture, and shape (Syazwani et al., 2022). The extracted features are subsequently used to train an image classifier capable of accurately distinguishing between healthy and diseased wheat. (Khan et al., 2023a) developed an automatic classification framework for wheat diseases based on machine learning techniques, effectively identifying wheat brown rust and yellow rust. (Bao et al., 2021b) presented an approach for detecting leaf diseases and their severity based on E-MMC metric learning, focusing on wheat mildew and stripe rust. However, in machine learning-based algorithms for identifying crop leaf pests and diseases, traditional image processing techniques or manually designed feature-based classification and recognition algorithms are commonly employed (Zhang et al., 2023). These algorithms are typically limited to extracting low-level features and struggle to capture deep and complex image information, failing to fully capture the complexity of sample data, which affects the accuracy of diagnosing localized regions of leaf diseases (Xu et al., 2024).

Recently, deep learning has made significant advancements in the field of crop pest and disease identification, achieving remarkable success in domains such as image processing (Chen et al., 2021b), natural language processing (Zeng and Xiong, 2022), and speech recognition (Kim et al., 2024), owing to its powerful representational capabilities (Zhang et al., 2024; Khan et al., 2023b). Advanced deep learning techniques, such as convolutional neural networks (CNNs) and attention mechanisms, have been applied to crop pest and disease detection (Jia et al., 2023). These methods can automatically, efficiently, and accurately extract target features from large datasets of crop leaf pest and disease images, thereby replacing traditional recognition approaches that rely on manual feature extraction. To facilitate rapid and accurate identification of wheat leaf diseases and reduce agricultural losses, (Jiang et al., 2021) introduced an enhanced VGG16 model integrated with a multi-task transfer learning strategy for detecting wheat leaf diseases. They modified the VGG16 model and employed a pre-trained model on the ImageNET platform for transfer learning and interactive learning. Experimental results demonstrated that this method outperformed single-task models, the ResNet50 model, and the DenseNet121 model. (Dong et al., 2024) presented the SC-ConvNeXt model for wheat disease identification. This network model utilizes ConvNeXt-T for feature extraction and incorporates an enhanced CBAM mechanism to mitigate the effects of interference from complex environmental factors. To improve the accuracy of a single category of wheat disease identification, (Nigam et al., 2023) focused solely on wheat rust and fine-tuned the EfficientNet B4 model for wheat disease recognition. (Chang et al., 2024) proposed the Imp-DenseNet model for identifying the three types of wheat rust, aiming to facilitate wheat rust identification in field environments. (Hassan et al., 2024) advanced the UNET detection model for yellow rust disease detection in wheat, achieving high classification accuracy for wheat diseases.

Deep learning constructs multi-layer neural network models that enable advanced data representation and understanding through hierarchical feature extraction and abstraction. However, as these multi-layer networks become deeper, each layer introduces numerous parameters and nonlinear activation functions (Chang, 2025). Although such architectures excel in handling complex data and tasks, their high complexity and nonlinearity lead to low transparency and poor interpretability (Goethals et al., 2022). Users often struggle to intuitively understand the logical basis behind model decisions, casting doubt on their credibility and perceiving deep models as data-driven "black box" systems (Marcus and Teuwen, 2024). The decision-making process in such models inherently involves high-dimensional nonlinear mappings, with internal reasoning mechanisms that lack explicit interpretability. This fundamentally complicates result attribution and causal inference. In agricultural applications, such as leaf disease identification, researchers have proposed various interpretability methods. These techniques such as feature visualization, attention mechanism analysis, and decision rule extraction (Hernández et al., 2024) aim to unveil the internal reasoning pathways of deep models during disease diagnosis, thereby enhancing model transparency and credibility.

However, most existing intrinsically interpretable models rely on spatially rigid prototypes, which are unable to explicitly explain the geometric changes in disease patterns and complex background feature information. This limitation restricts the provision of detailed explanations and improved recognition accuracy (Ma et al., 2024). Therefore, in this work, we propose an interpretable wheat leaf disease identification model (CDPNet) based on a deformable prototypical part network and contrastive learning. In CDPNet, each prototype comprises multiple prototypical parts that adaptively adjust their spatial positions relative to one another depending on the input image. This allows each prototype to detect object features with greater tolerance of spatial transformations, since the parts within a prototype can move. To identify wheat leaf disease types and uncover the infected regions influencing model decisions, we first employ Deformable ProtoPNet (Donnelly et al., 2022) to calculate the similarity values relating to the convolutional feature maps of the image and the latent prototype features. Generally, a higher similarity

score indicates a greater influence of that region on the model's decision. Secondly, to effectively distinguish the target regions of wheat leaf diseases from complex backgrounds and enhance the model's feature extraction capabilities, we introduce the CA Mechanism (Lin et al., 2022; Chen et al., 2021a). This mechanism guides the model to focus on spatial contextual features. By amplifying differences between disease areas and surrounding backgrounds, it significantly enhances the discriminative power of disease features, thereby improving recognition performance in complex scenarios. Finally, in practical applications, some wheat leaf diseases exhibit low incidence rates and high image acquisition costs, leading to limited training data. To address this challenge, we introduce the self-supervised contrastive learning strategy Barlow Twins (Zbontar et al., 2021). This approach maximizes similarity between different transformed versions of the same image while minimizing similarity between distinct images, thereby enabling deep exploration of discriminative features across wheat leaf disease instances. In summary, the main contributions of this work are summarized as follows:

- The deformable prototype network in CDPNet is designed to adaptively adjust relative spatial positions through flexible and dynamic prototype learning, thereby providing clinical interpretability for the identification of wheat leaf diseases.
- We propose a novel interpretable model for wheat leaf disease identification—the Contrastive Deformable Prototypical part Network (CDPNet). This model is capable of discovering key regions in wheat leaf disease images that influence the model's decisions. Additionally, it effectively distinguishes between disease target regions and complex backgrounds, and deeply mines latent feature information among samples, offering a more comprehensive and in-depth analytical perspective for disease identification.
- We have created a real-world wheat leaf disease dataset to facilitate further research on disease identification in practical field environments.
- Through extensive experimentation using the wheat leaf disease dataset, as well as other public crop disease datasets, the results demonstrate that CDPNet achieves superior identification performance, outperforming classical models, and validating its generalization ability and interpretability.

2 Related work

2.1 Leaf disease identification based on machine learning

The recognition of crop leaf diseases has long been a central research focus within the field of agricultural engineering (Thakur et al., 2022). The application of modern information technologies for diagnosing and identifying crop leaf diseases provides an advanced, systematic, and effective approach (Balakrishna and Rao, 2019). Research on leaf disease identification methods can be broadly categorized into two primary approaches: traditional machine learning techniques and contemporary deep learning approaches.

Machine learning is utilized to automatically analyze large-scale datasets, uncover latent patterns, and apply these insights to subsequent analysis and prediction tasks. With the advancement of image processing technologies, machine learning has been extensively applied to leaf disease identification (Thakur et al., 2022). Researchers employ feature extraction and segmentation techniques to capture key disease characteristics, which are subsequently classified using machine learning algorithms. Under conditions of limited computational resources, machine learning initially emerged as the primary research tool, producing notable results. (Balakrishna and Rao (2019) conducted experiments on tomato leaf diseases, initially categorizing tomato leaves into healthy and diseased classes using the K-Nearest Neighbors (KNN) method, followed by effective sub-classification of diseased leaves using a combination of Probabilistic Neural Networks (PNN) and KNN. (Pattnaik and Parvathi, 2021) utilized the Histogram of Oriented Gradients (HOG) to characterize features extracted from segmented images, which were then input into a Support Vector Machine (SVM) for classification. Due to the relatively low classification difficulty, their test accuracy reached 97%. (Javidan et al., 2023) utilized K-means clustering technology to locate infected regions in images and subsequently accomplished grape leaf disease classification through SVM. However, machine learning-based approaches to leaf disease recognition, while capable of distinguishing certain disease features and generating classification results, continue to exhibit several limitations (Wani et al., 2022): (1) Feature selection limitations: Traditional machine learning approaches require the manual selection of features to describe pest or disease images. However, such features often capture only partial image information. Moreover, the variability of pests and diseases across growing environments renders selected features insufficient to comprehensively represent all relevant characteristics. (2) Feature extraction challenges: Machine learning cannot automatically extract features, necessitating manual extraction, which is also highly sensitive to image noise. (3) Limited generalizability and recognition scope: Trained models can typically recognize only the specific crop pests and diseases on which they were trained, making it difficult to extend recognition capabilities to other disease types. (4) Narrow application scope: Constrained by disease-specific characteristics, these methods are generally limited to learning and classifying features of particular crops in specific regions, which restricts their applicability across a broader range of species.

Compared with traditional machine learning methods, deep learning addresses inefficiencies and low accuracy arising from manually designed features in complex environments. In recent years, alongside the ascent of deep learning advancements, CNNs and Transformers have undergone rapid development (Khan et al., 2023a). The convolutional layers of CNNs utilize a local receptive

field design, in which each neuron is connected only to a restricted region of the input image (Quan et al., 2022). This design is well-suited to image data, since local information (e.g., edges, textures) plays a critical role in object recognition (Xu et al., 2024). (Bao et al., 2022) presented an enhanced recognition network called AX-RetinaNet. This model employs an X-module enhanced multi-scale feature integration and channel attention for feature extraction, thereby enabling effective detection and classification of tea diseases, with an identification accuracy reaching 96.75%. To address the issue of abnormal recognition caused by various image distortions in the healthy and diseased parts of coffee plant leaves, (Nawaz et al., 2024) suggested a CoffeeNet model. The model under consideration makes use of a ResNet-50 framework and an attention mechanism for the purpose of extracting features of diverse coffee leaf diseases. To increase the accuracy of classifying plant leaf diseases while keeping the model lightweight, (Zhao et al., 2024) developed a neural architecture termed CAST-Net. This lightweight network model is based on a combination of convolution and self-attention. It further employs a self-distillation method to enhance the precision of leaf disease classification while reducing model parameters and failure cases. The findings indicate that, in comparison with existing models, CAST-Net attains enhanced precision, reduced parameter complexity, decreased training time, and lower computational complexity. The Transformer architecture captures global dependencies among elements of input sequences (Khan et al., 2022b). In image classification, the self-attention mechanism allows the model to incorporate information from all pixels or features when processing each individual one (Xu et al., 2021). This enables Transformers to more effectively capture the overall structure and contextual information of images, providing advantages for classification tasks that rely on global information. (Borhani et al., 2022) proposed a lightweight model based on Vision Transformer for plant disease classification. To better 174 leverage the strengths of both CNNs and Transformers, (Alshammari et al., 2022) utilized a deep ensemble learning strategy to combine a CNN with a vision transformer model for the purpose of classifying Olive Diseases. (Thakur et al., 2023) also proposed a composite model that integrates the advantages of ViT with the innate feature extraction capabilities of CNNS for plant leaf disease recognition.

2.2 Interpretable leaf disease classification using deep neural networks

Image classification, a fundamental task in computer vision, focuses on achieving accurate multi-class categorization based on image content while minimizing error. Machine learning initially demonstrated significant potential in image classification, and within this domain, deep learning gradually emerged as the more suitable approach. CNNs, characterized by local connectivity and translation invariance, align well with the inherent properties of image data. Despite continual improvements in classification accuracy, researchers have identified persistent challenges in deep learning for image tasks, including adversarial robustness, generalization, and fairness. Interpretability research provides a critical pathway to address the "black box" nature of deep learning (Zhang et al., 2025). Its objective is to elucidate model decision-making mechanisms through human-understandable methods, thereby enhancing credibility and robustness. From a modeling perspective, Interpretability research can be broadly categorized into two types: *post-hoc* interpretation methods and intrinsically interpretable models.

1. *Post-hoc* interpretation methods. These primarily target black-box models, analyzing them through various algorithms such as visualization analysis, importance analysis, etc., to infer the model's decision-making procedure. Examples include Feature Attribution, Permutation Importance, and Class Activation Mapping (CAM). For instance, (Mishra et al., 2024) proposed an image-based interpretable leaf disease detection framework (I-LDD) that utilizes Local Interpretable Model-agnostic Explanations (LIME) to obtain explanations for model classifications. Similarly, (Raval and Chaki, 2024) employed LIME technology, taking leaf diseases as an example, and (Chakrabarty et al., 2024) used interpretable artificial intelligence to visualize the decision-making processes of their model, focusing on rice leaf diseases. To offer a more thorough understanding of the model's interpretability, (Hernández et al., 2024) adopted the Grad-CAM method to visualize the infected regions of grape leaves, explaining the neural network's attribution to leaf disease detection. (Wei et al., 2022) presented the ResNet-CBAM model for interpretable leaf disease classification and compared three visualization methods: SmoothGrad, LIME, and GradCAM, to conduct *post-hoc* interpretability of the model. Meanwhile, (Dai et al., 2024) employed t-SNE and SHAP visualization methods to explain whether the model focuses on plant pest and disease characteristics.
2. Intrinsically interpretable models. Intrinsically interpretable models require us to select human-understandable features and adopt models with good interpretability during the problem-solving process (Jiang et al., 2025). This objective is realized through the construction of models that are self-explanatory and which incorporate interpretability directly into their structures. Such models include decision trees, rule-based models, linear models, and attention models. Our model belongs to the category of intrinsically interpretable models, which integrate interpretability into the specific model structure, enabling the model itself to possess interpretability. The model outputs not only the results but also the reasons behind those results, thereby ensuring the reliability and safety of the interpretations. CDPNet discovers key regions influencing model decisions and predicts pest and disease categories by computing similarity of the convolutional feature maps of images to the latent prototype features, thus explaining the model's decision-making process and attribution. Through flexible

and dynamic prototype learning, it achieves accurate identification of wheat leaf diseases in natural field environments along with rich interpretability.

3 Materials and methods

3.1 Dataset acquisition and image preprocessing

This study utilized a hybrid data source to construct a wheat leaf disease dataset. The self-constructed dataset was compiled by the research team under expert guidance through field photography conducted in Fengyang County, Chuzhou City, Anhui Province, from April 15 to May 15, 2024. Fieldwork was conducted daily between 8:00 AM and 6:00 PM. Images were captured using a Vivo Y70s smartphone, covering six common wheat leaf diseases: Brown Rust, Healthy, Leaf Blight, Mildew, Septoria, and Yellow Rust. A total of 1,340 valid images were obtained. Figure 1 illustrates images of wheat leaf diseases from various categories.

To enhance the dataset, this study also incorporated wheat leaf disease images from the Wheat Plant Diseases dataset on Kaggle. This dataset is designed to enable researchers and developers to build robust machine learning models for classifying various wheat plant diseases. It provides a collection of high resolution images depicting real-world wheat diseases without relying on artificial augmentation techniques. Data filtering was performed on this dataset to remove duplicate and misclassified images from the original public dataset. This process resulted in the creation of a wheat leaf disease dataset (WL-Disease) comprising six categories and a total of 6,513 images. The specific categories and their corresponding image counts are detailed in Table 1.

In the WL-Disease dataset, all training images are labeled without annotations on specific image regions. The dataset was randomly divided into training and testing sets at an 80:20 ratio to ensure the validity and fairness of model training and validation.

To facilitate model training, all disease images were uniformly resized to 500 × 500 pixels and converted to JPG format. Data augmentation techniques enhance the effectiveness of neural networks by increasing both the heterogeneity and volume of training data, thereby improving generalization capabilities. Throughout the experiment, due to the limited number of samples per class in the dataset, we applied 10-fold offline data augmentation to mitigate overfitting to specific subsets and improve the model's stability and accuracy in practical applications. This process included random rotation, 45-degree skew, 10-degree shear operations, 5-strength distortion processing, 50% probability of left-right flip, and color enhancement to expand the training set. Figure 2 shows the comparison before and after image augmentation.

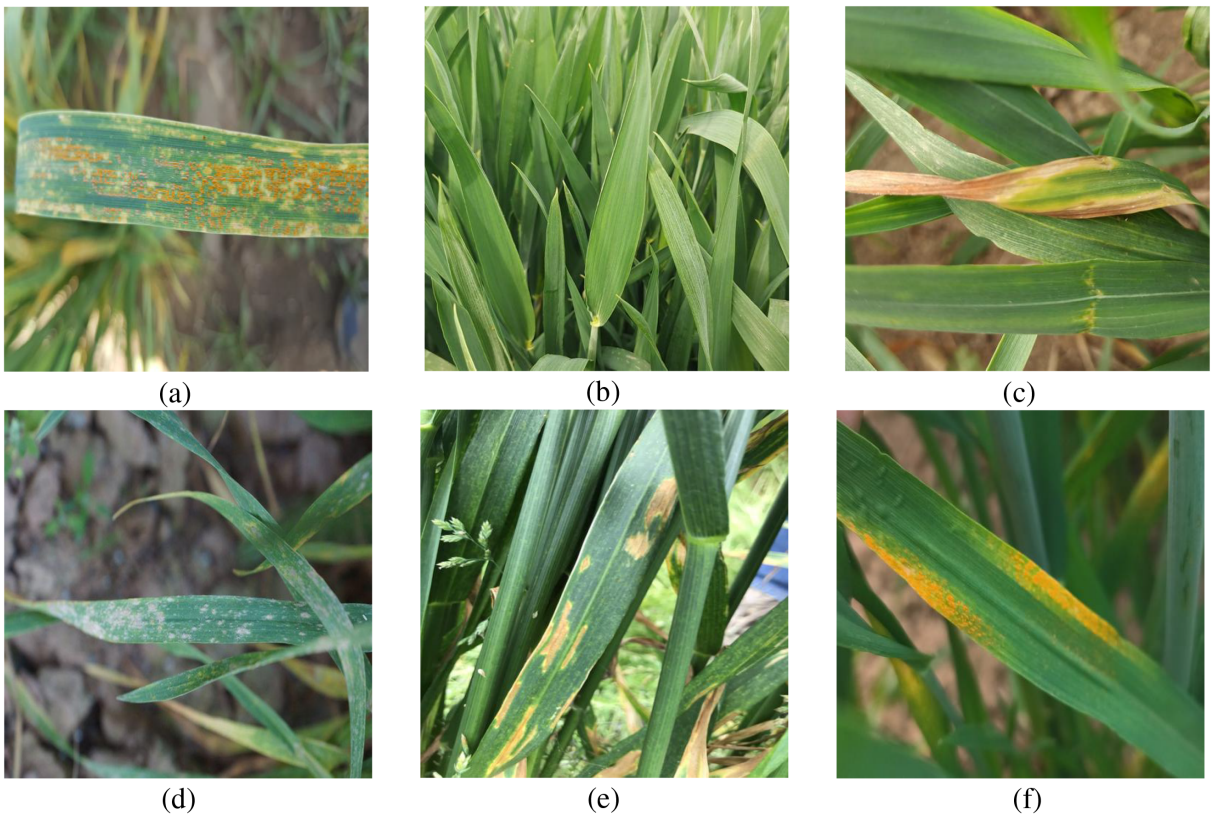

FIGURE 1
Samples of wheat leaf disease. **(a)** Brown Rust, **(b)** Healthy, **(c)** Leaf Blight, **(d)** Mildew, **(e)** Septoria, **(f)** Yellow Rust.

TABLE 1 Detailed descriptions of the various types of samples within the WL-disease dataset.

Category	Number	Train set	Test set
Brown Rust	1054	843	211
Healthy	812	645	167
Leaf Blight	1008	806	202
Mildew	1328	1062	266
Septoria	916	732	184
Yellow Rust	1395	1116	279
Total number	6513	5204	1309

3.2 Problem formulation

Currently, the task of wheat leaf disease identification aims to assign the correct label from a predefined set of categories to an image, achieving precise classification and recognition. A common research approach involves utilizing deep learning algorithms to extract features of wheat leaf diseases and perform recognition. In contrast, this study adopts a methodology that incorporates a deformable prototypical part network with contrastive learning, aiming to achieve interpretable and accurate recognition of wheat leaf diseases. Given a leaf disease image x, its corresponding category label is $y \in \{0,\ldots,c,\ldots,C\}$. The model learns a mapping function $\mathcal{F}:\mathcal{F}(x) \rightarrow \hat{y}$ capable of predicting the category to which the given image x belongs, where $\hat{y}$ is the probability that the wheat leaf disease image x belongs to its corresponding category. The objective of this research is to optimize the mapping function $\mathcal{F}$ to maximize the predicted probability. Meanwhile, the method automatically identifies the affected regions of wheat leaf diseases, providing interpretable evidence for the final classification results.

3.3 CDPNet network architecture

In this section, we provide a detailed description of the architecture of the proposed interpretable wheat leaf disease recognition model based on a deformable prototypical part network and contrastive learning, which is visualized in Figure 3. CDPNet primarily consists of convolutional layers f, a deformable prototype layer $\mathcal{G}$, and a fully connected last layer h. Given an input image $x \in X$, the convolutional layers f first extract a meaningful image representation $\mathbf{Z} = f(x) \in \mathbb{R}^{H\times W\times C}$ (with height H, width W, and number of channels C). Second, for each prototype, the deformable prototype layer $\mathcal{G}$ computes a similarity matrix $\mathbf{M}^{x}_{p_i} \in \mathbb{R}^{H\times W}$ between the convolutional feature maps $\mathbf{Z}$ and a learnable latent prototype feature representation $\mathbf{P}^{(c,t)} \in \mathbb{R}^{1\times 1\times C}$ (the t-th prototype of class c). The similarity maps contain positive scores indicating where and to what extent prototypes are present in an image. CDPNet uses the highest value of the similarity map as the final similarity score between $\mathbf{P}^{(c,t)}$ and x, indicating how strong the prototype $\mathbf{P}^{(c,t)}$ is present in x. Finally, the similarity scores from the deformable prototype layer $\mathcal{G}$ are aggregated in the fully connected layer h to generate the final classification logits. These logits are normalized using the softmax function to obtain the predicted probability distribution of disease categories. In addition, to facilitate the visualization of prototypes as specific prototypical parts of a sample, the learned prototypes are substituted with the closest feature representation from authentic training images, thereby ensuring interpretability.

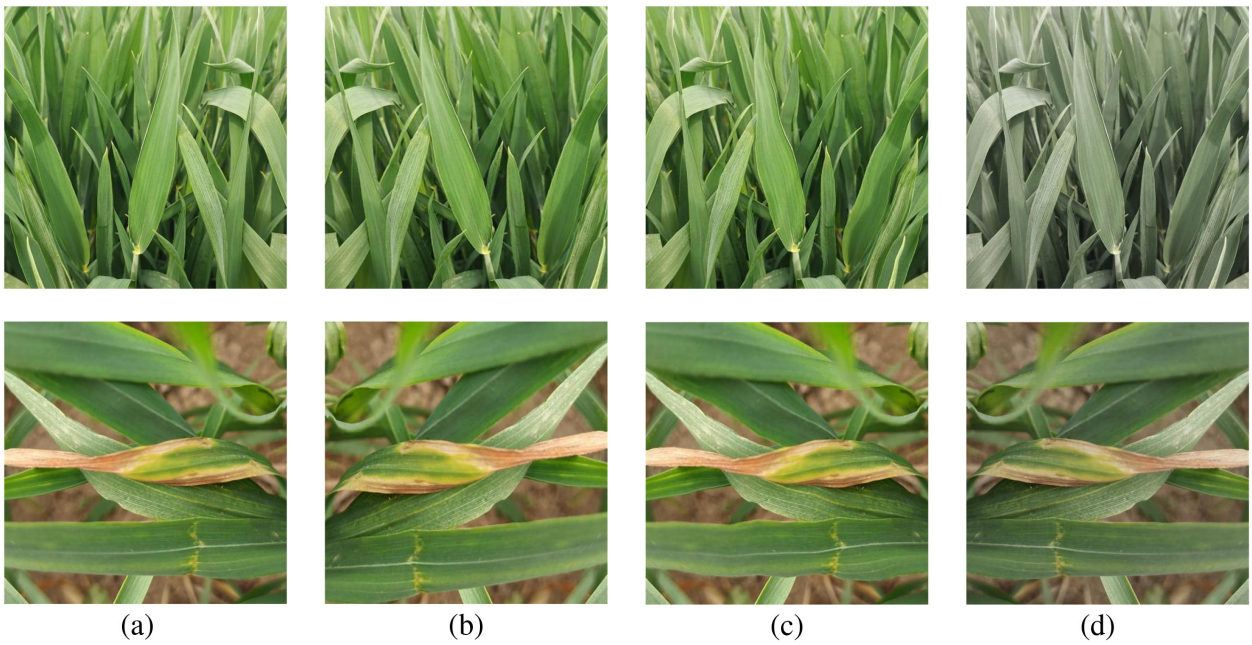

FIGURE 2
Samples of image augmentation of WL-Disease dataset. **(a)** Original image, **(b)** Left-right flip, **(c)** Distortion, **(d)** Color enhancement.

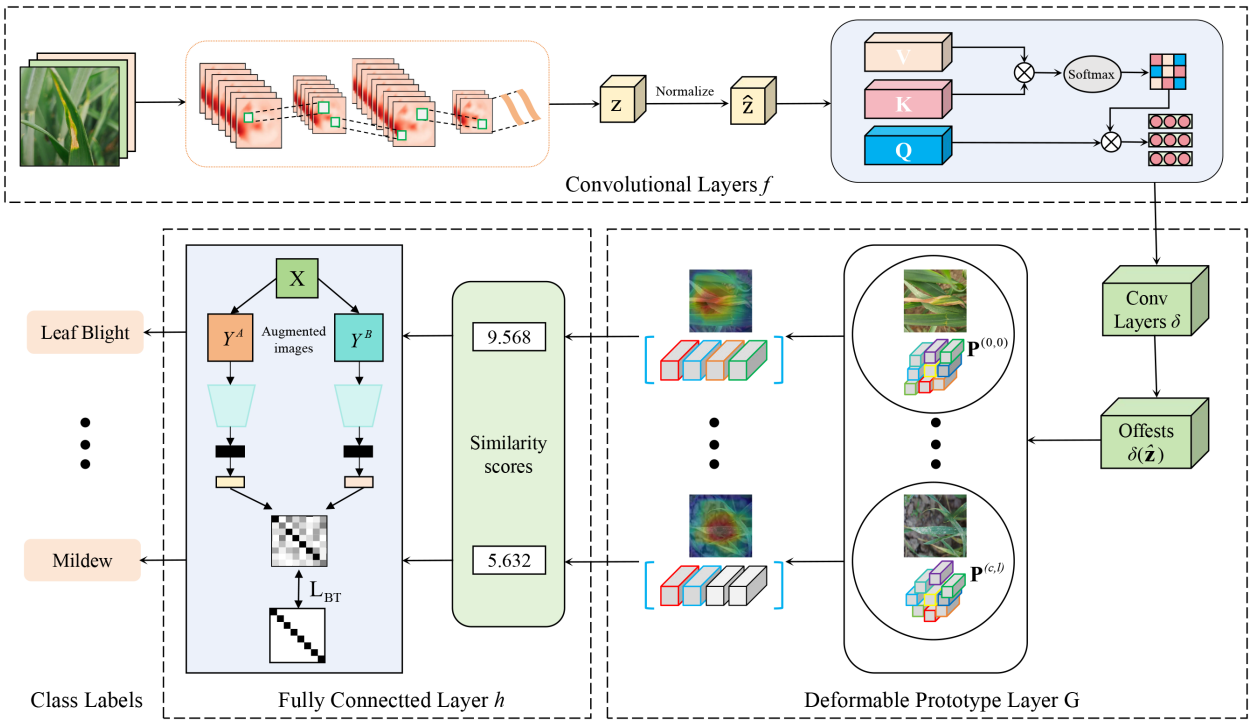

FIGURE 3
Visualization of the architecture of the proposed CDPNet.

3.3.1 Convolutional layer

The role of the convolutional layers extract information from the input image, which is referred to as image features. These features are manifested through combinations or individual contributions of each pixel within the image, such as texture and color characteristics. Through the convolutional layers, local regional feature extraction of wheat leaf disease images can be achieved, generating the original feature representation of the image. Specifically, the convolutional layers f borrow the convolutional layers from classical models (such as VGG19, ResNet152, DenseNet161, etc.), and then two additional 1×1 convolutional layers intended to modify the number of channels present in the top-level feature maps. Meanwhile, we use ReLU as the activation function for all convolutional layers, except for the last layer, which employs the sigmoid activation function. Equation 1 converts the input image x into a feature vector.

$$[\mathbf{z}_1, \ldots, \mathbf{z}_i, \ldots, \mathbf{z}_m] = \text{Conv2D}(\text{x}) \in \mathbb{R}^{W \times H \times C} \tag{1}$$

To effectively distinguish the target regions of wheat leaf diseases from complex backgrounds, our core method is to employ a CA mechanism, as shown in Figure 4. CA mechanism enables the model to dynamically construct cross-modal feature correlation matrices, allowing it to adaptively focus on key discriminative features such as lesion textures and color distortions. It also facilitates a more comprehensive integration of contextual information from multiple sources, consequently boosting both the precision and the generalization performance of the recognition task. Firstly, the correlation scores indicating the similarity between the query and keys are determined by calculating the dot product of the query $\mathbf{Q}$ and keys $\mathbf{K}$. Secondly, these similarities are transformed into a probability distribution using the softmax function, representing the attention weights of the query with respect to each key. These attention weights are then applied to the values $\mathbf{V}$, ultimately resulting in the output vector. Mathematically, the formula for cross-attention is presented in Equation 2:

$$\text{CrossAttention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\text{T}}}{\sqrt{d_k}}\right) \tag{2}$$

where, $\mathbf{Q}\mathbf{K}^T$ represents the dot product of the query and the key, indicating the similarity between the two sequences at different positions; d_k is the dimension of the key, which serves as a scaling factor to prevent excessively large numerical values.

3.3.2 Deformable prototype layer

The fundamental idea behind the deformable prototype layer $\mathcal{G}$ is to find highly interpretable (i.e., representative) deformable prototypes by calculating the similarity scores s between the convolutional feature maps $\mathbf{Z}$ of a test image x and the prototypes $\mathbf{P}$. Each part of these prototypes corresponds to key regions that influence the model's decision-making processes, and these regions could be visualized. For a CDPNet, the L^2-length of all prototype parts $\mathbf{P}_{m,n}^{(c,t)}$ of all deformable prototypes $\mathbf{P}^{(c,t)}$ is the same. Furthermore, at the spatial location (a,b) of each image feature tensor $\hat{\mathbf{z}}$, the corresponding vectors also possess are of equal L^2-length, as shown in Equations 3 and 4.

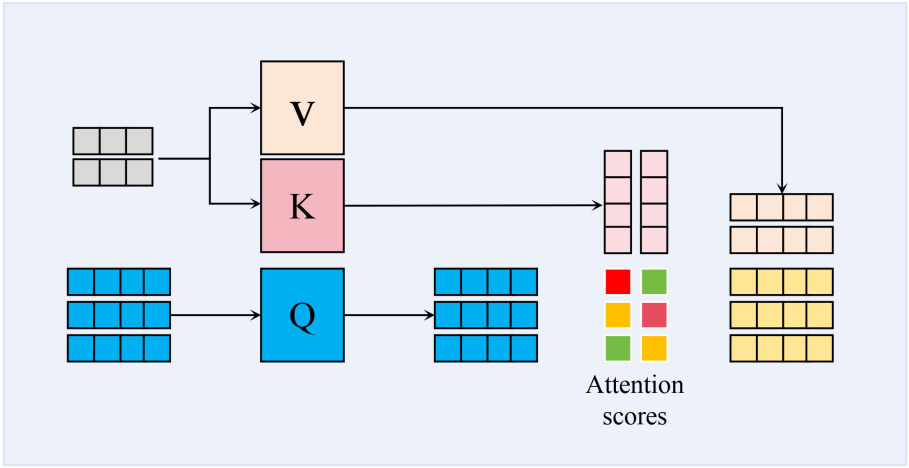

FIGURE 4
Schematic representation of cross attention mechanism.

$$\| \mathbf{P}_{m,n}^{(c,t)} \|_2 = r = \frac{1}{\sqrt{\rho}}, \tag{3}$$

$$\| \hat{\mathbf{z}}_{a,b} \|_2 = r = \frac{1}{\sqrt{\rho}} \tag{4}$$

Then, the formula for calculating the similarity of deformable prototypes $\mathbf{P}^{(c,t)}$ and the image feature tensor $\hat{\mathbf{z}}$ defined as shown in Equation 5.

$$\mathcal{G}(\hat{\mathbf{z}})_{a,b}^{(c,t)} = \sum_m \sum_n \mathbf{P}_{m,n}^{(c,t)} \cdot \hat{\mathbf{z}}_{a+m,b+n} \tag{5}$$

In order to facilitate the deformation of a deformable prototype $\mathbf{P}^{(c,t)}$, it has been proposed that offsets δ (2D vector) be introduced, thereby enabling each constituent part $\mathbf{P}_{m,n}^{(c,t)}$ of the prototype to migrate in relation to the spatial location (a, b) with respect to the image feature tensor $\hat{\mathbf{z}}$ when the prototype is applied. Mathematically, the formula for calculating the similarity of the prototype is defined as shown in Equation 6.

$$\mathcal{G}(\hat{\mathbf{z}})_{a,b}^{(c,t)} = \sum_m \sum_n \mathbf{P}_{m,n}^{(c,t)} \cdot \hat{\mathbf{z}}_{a+m+\Delta_1,b+n+\Delta_2} \tag{6}$$

The maximum similarity with respect to an arbitrary set of positions is given by the following definition.

$$\mathcal{G}(\hat{\mathbf{z}})^{(c,t)} = \max_{a,b} \mathcal{G}(\hat{\mathbf{z}})_{a,b}^{(c,t)} \tag{7}$$

Figure 5 shows the operational process of the deformable prototypes. The input $\hat{\mathbf{z}}$ undergoes processing by the offset prediction function δ, resulting in (b) a grid of offset values. Subsequently, these offsets are utilized to (c) modify the spatial positions of each prototypical part. After this adjustment, the updated prototypical parts are (d) aligned with the input to (e) compute the prototype similarity in accordance with Equation 6.

3.3.3 Fully connected layer

The fully connected layer integrates and abstracts the features learned from the preceding layers to facilitate the execution of classification or regression tasks. It performs a linear transformation on the input data using a weight matrix and a bias vector. In the CDPNet model, the fully connected layer multiplies the similarity scores generated by the deformable prototype layer by the weight matrix $\mathbf{W}$ in the fully connected layer. The result is then feeds the result into the Softmax layer for normalization. Finally, it generates a prediction result for the given leaf disease and pest image. The prediction of the leaf disease image at this point is calculated as shown in Equation 8.

$$\hat{y} = \text{Softmax}(\mathbf{R}_{zg}\mathbf{W}_\text{h} + \text{b}) \tag{8}$$

where, $\mathbf{W}_h \in \mathbb{R}^{d \times c}$ is the parameter matrix, represents the image features, b is the bias term, and $\hat{\mathbf{y}} = [\hat{y}_0, \ldots, \hat{y}_c, \ldots, \hat{y}_C]$, $\hat{y}_c$ denotes the predicted probability that the input image belongs to the c-th class. Therefore, given an image x, a novel form of cross-entropy is employed: the margin-subtracted cross-entropy. The formula is shown in Equation 9.

$$\mathcal{C}_{ce}(\theta) = \sum_{i=1}^{N} -\log \frac{\exp\left(\sum_{c,t} \mathbf{W}_h^{((c,t),y^{(i)})} \mathcal{G}^{(-)}(i)^{(c,t)}\right)}{\sum_{c'} \exp\left(\sum_{c,t} \mathbf{W}_h^{((c,t),c')} \mathcal{G}^{(-)}(i)^{(c,t)}\right)} \tag{9}$$

where, θ represents the parameters that need to be learned, and $\mathbf{W}_h^{((c,t),c')}$ denotes the connections between the deformable prototypes $\mathbf{P}^{(c,t)}$ and the last layer responsible for computing similarity with the c' classes.

3.3.4 Model learning

As deep learning progresses, approaches for identifying wheat leaf diseases harness deep networks to automatically learn features;

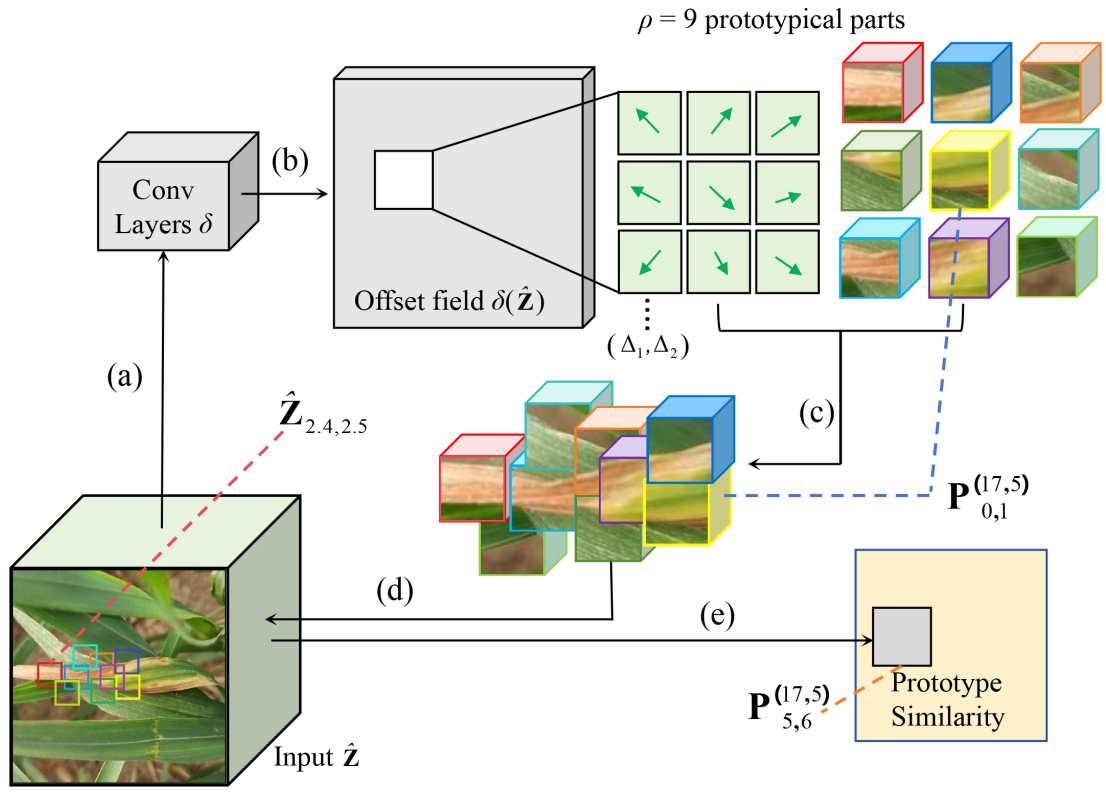

FIGURE 5
CPNet is applied to the latent representation of the Leaf Blight. **(a)** Put the input features into the offset prediction function to generate **(b)** an offset field. Then, **(c)** adjust the spatial position of each prototypical part using these offsets, **(d)** compare the adjusted parts with the input, and **(e)** calculate the prototype similarity.

however, these methods heavily depend on the availability of a substantial volume of training data. To address the limited availability of wheat image data, we have introduced a self-supervised contrastive learning approach to tackle the challenge of recognition with limited samples. Specifically, Figure 6 shows that we used the Barlow Twins in contrastive learning to conduct feature learning between samples. Barlow Twins represents a self-supervised learning approach for representation learning, stemming

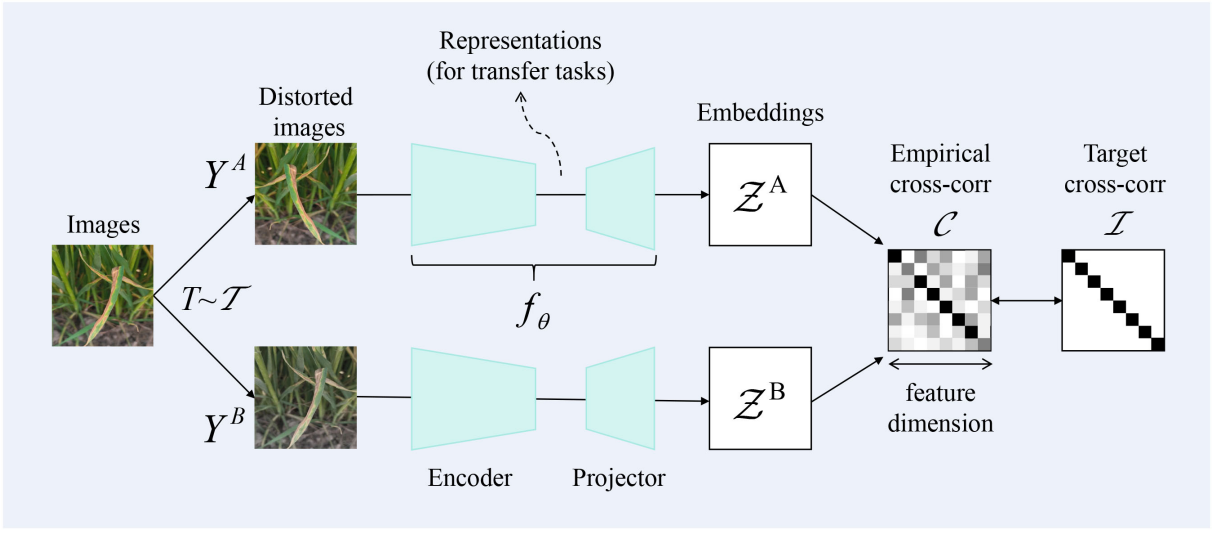

FIGURE 6
Schematic representation of Barlow Twins.

from the groundbreaking ideas of the JPT team. Its core lies in minimizing the covariance distance between twin networks, enabling their learned features to be as independent as possible while maintaining similarity. This approach not only enhances the efficiency of the model but also achieves favorable pre-training results even with scarce data.

Barlow Twins is a self-supervised learning method rooted in information theory, with the objective of reducing redundancy among neurons. This approach mandates that neurons remain invariant to data augmentations while being independent of one another. During actual training, the parameters of the neural network are adjusted through backpropagation to maximize the diagonal elements of the cross-correlation matrix and minimize the off-diagonal elements — approaching an identity matrix — thereby achieving the aforementioned goal. It is calculated as shown in Equation 10.

$$\mathcal{L}_{BT} = \sum_i (1 - C_{ii})^2 + \lambda \sum_i \sum_{j \neq i} C_{ij}^2 \tag{10}$$

where λ is a positive constant trading off the importance of the first and second terms of the loss, $\sum_i (1 - C_{ii})^2$ is an invariance term (diagonal or identity term) designed to direct neurons to produce the same output under different augmentations, $\sum_i \sum_{j \neq i} C_{ij}^2$ is a redundancy reduction term (off-diagonal term) intended to make each neuron produce a different output.

$$C_{ij} = \frac{\sum_b \mathcal{Z}_{b,i}^A \mathcal{Z}_{b,j}^B}{\sqrt{\sum_b \left(\mathcal{Z}_{b,i}^A\right)^2} \sqrt{\sum_b \left(\mathcal{Z}_{b,j}^B\right)^2}} \tag{11}$$

where, b denotes the index of the batch, while i and j represent the feature dimensions of the network's output (i.e., they correspond to the values in the i-th and j-th dimensions of two vectors within the current batch). C_{ij} is the element value at the i-th row and j-th column of matrix C. It is equal to the sum of the products of the i-th dimension of the augmented feature vector $\mathcal{Z}^{\mathrm{A}}$ and the j-th dimension of the augmented feature vector $\mathcal{Z}^{\mathrm{B}}$ for different pairs within the batch. The summation is primarily carried out over the current batch size. Matrix C is a square matrix, and its dimensions correspond to the output dimension of the network (assuming each embedding dimension output by the network is D, then the dimensions of square matrix C are D × D). The values of matrix C range between -1 (indicating perfect negative correlation) and 1 (indicating perfect positive correlation).

In order to discover a meaningful feature space in which the image features belonging to class c are found to cluster around the prototypes of the same class while being segregated from features of other classes within a hypersphere, CDPNet employs Stochastic Gradient Descent (SGD) to perform optimization on the features of the convolutional layer f and the deformable prototype layer $\mathcal{G}$. In this process, SGD incorporates both cluster and separation losses and adjusts the angular space. These two losses are defined as shown in Equations 11 and 12.

$$\mathcal{C}_{clst} = -\frac{1}{N} \sum_{i=1}^{N} \max_{\mathbf{P}^{(c,t)} : c = y^{(i)}} \mathcal{G}(\hat{\mathbf{z}}^{(i)})^{(c,t)} \tag{12}$$

$$S_{\text{sep}} = \frac{1}{N} \sum_{i=1}^{N} \max_{\mathbf{P}^{(c,t)} : c \neq y^{(i)}} \mathcal{G}(\hat{z}^{i})^{(c,t)} \tag{13}$$

where, N represents the total number of inputs, $\hat{\mathbf{z}}^{(i)}$ denotes the normalized and scaled image feature tensor of input i at each spatial location, $\mathrm{y}^{(i)}$ is the label corresponding to input $\mathrm{x}^{(i)}$, and all other values are consistent with the definitions provided in the preceding context.

Although the subtraction margin encourages separation among categories, it does not promote diversity among intra-class prototypes or within prototype parts within a prototype. Specifically, deformations without further regularization often lead to redundancy among prototype parts within a prototype. To mitigate this issue, we prevent this behavior by introducing an orthogonality loss among prototype parts. Its formula is shown in Equation 14.

$$O_{\text{ortho}} = \sum_c \left\| \mathbf{P}^{(c)} \mathbf{P}^{(c)\top} - r^2 \mathbf{I}^{(\rho L)} \right\|_F^2 \tag{14}$$

where L is the number of deformable prototypes in class c, ρL represents the total number of prototype parts across all prototypes in class c, $\mathbf{I}^{(\rho L)}$ is the $\rho L \times \rho L$ identity matrix, and $\mathbf{P}^{(c)} \in \mathbb{R}^{\rho L \times d}$ is a matrix where each prototype part of every prototype in class c is arranged as a row.

Finally, the overall loss function during the CDPNet training process is formulated as shown in Equation 15.

$$L_{total} = \mathcal{C}_{ce}(\theta) + \lambda_1 \mathcal{C}_{clst} + \lambda_2 \mathcal{S}_{\text{sep}} + \lambda_3 O_{\text{ortho}} + \lambda_4 \mathcal{L}_{BT} \tag{15}$$

4 Results and analysis

4.1 Experimental setup

In this study, the PyTorch framework was utilized. PyTorch is an open-source library designed for deep learning tasks, offering a concise, elegant, efficient, and rapid framework that serves as a deep

TABLE 2 Test system environment configuration.

System environment	Configuration
Operating system	Ubuntu 18.04
GPU	V100-32GB(32GB)
CPU	10 vCPU Intel Xeon Processor (Skylake, IBRS)
Pytorch	PyTorch 1.8.0
Python	Python 3.8
Batch size	32
Epoch	50

learning research platform providing maximum flexibility and speed. The experimental environment and parameters used in this study are detailed in Table 2.

4.2 Evaluation metrics

We validated the model's effectiveness on the test set using standard classification performance metrics. These metrics include accuracy, precision, recall, F1-score, and AUC. Their mathematical expressions are as shown in Equations 16–21. All samples were categorized into four groups based on the differences between the true and predicted classes: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 - Score = 2 \star \frac{Precision \star Recall}{Precision + Recall} \tag{19}$$

$$TPR = \frac{TP}{TP + FN} \tag{20}$$

$$FPR = \frac{FP}{FP + TN} \tag{21}$$

In addition, we employed the confusion matrix and Receiver Operating Characteristic (ROC) curve to evaluate the model's performance. The confusion matrix and ROC curve indicate the model's credibility. The higher the ROC curve is positioned in the top-left corner, the better the model's performance. Meanwhile, we utilized CDPNet to visualize the prototype image classification activation maps and similarities, aiming to uncover the critical factors underlying the model's classification decisions and assist researchers in understanding the basis for the model's final classifications.

TABLE 3 Comparison of experimental results of different models on wheat leaf disease dataset.

No.	Data augmentation methods	Accuracy (%)
1	Resize(224,224)	92.25
2	Resize(224,224)+skew	92.56
3	Resize(224,224)+skew+shear	93.36
4	Resize(224,224)+skew+shear+distortion	93.82
5	Resize(224,224)+skew+shear+distortion+left-right flipping	94.76
6	Resize(224,224)+skew+shear+distortion+left-right flipping+color enhancement	95.83

4.3 Experimental results and comparative analysis

4.3.1 Performance evaluation of different data augmentation methods

Table 3 shows the results of experiments conducted using the CDPNet-DenseNet161 model with various data augmentation methods. Six distinct data augmentation schemes were generated by combining different techniques. Scheme 1 involved inputting the original image into the model after normalization (resizing to

TABLE 4 Comparison of experimental results of different models on wheat leaf disease dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC (%)
VGG19	90.38	90.67	90.89	90.50	97.69
ResNet152	91.61	91.36	91.08	91.03	97.92
DenseNet161	92.18	91.86	91.63	91.76	98.16
ProtoPNet-VGG19	91.85	91.38	91.45	91.35	98.02
ProtoPNe-ResNet152	92.16	91.82	91.52	91.66	98.08
ProtoPNet-DenseNet161	92.81	92.45	92.63	92.53	98.33
Deformer ProtoPNet-VGG19	92.15	91.83	91.52	91.65	98.12
Deformer ProtoPNe-ResNet152	92.63	92.27	92.11	92.19	98.25
Deformer ProtoPNet-DenseNet161	93.48	93.13	92.89	92.99	98.52
CDPNet-VGG19	94.22^c	93.72^c	93.97^c	93.77^c	99.16^c
CDPNet-ResNet152	94.89^c	94.21^c	94.47^c	94.29^c	99.38^c
CDPNet-DenseNet161	95.83^c	95.32^c	95.07^c	95.13^c	99.45^c

^{c}Denotes the test of statistical significance $p < 0.001$.

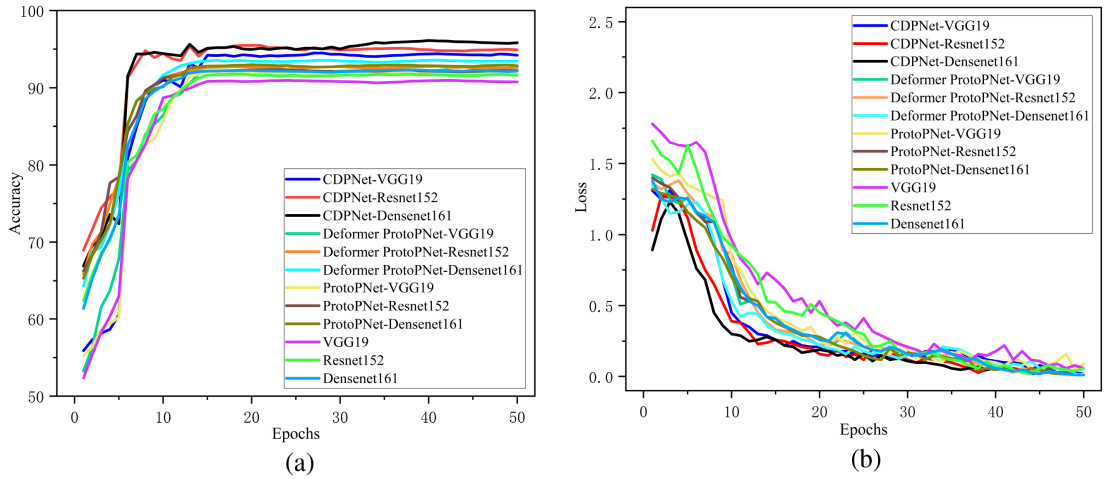

FIGURE 7
Comparison of the variation curves for loss values and accuracy across different models. **(a)** Accuracy curve, **(b)** Loss curve.

224×224×3), resulting in a classification accuracy of 92.25%. Subsequently, the introduction of various data augmentation methods, including skew, shear, distortion, left-right flipping, and color enhancement, to Scheme 1 led to an improvement in model accuracy. Among the augmentation techniques tested, color enhancement produced the most favorable results. The results indicate that Scheme 6 achieved the highest accuracy (95.83%), establishing it as the optimal data augmentation scheme.

4.3.2 Model performance comparisons

To validate the classification performance of the proposed CDPNet model for wheat leaf diseases, comparative experiments were conducted under identical conditions using the WL-Disease dataset, comparing CDPNet with VGG19, ResNet152, DenseNet161, ProtoPNet, and Deformer ProtoPNet models. The comparative results for each model are shown in Table 4. Figure 7 shows the loss value and accuracy comparison curves during the training phase for different model. Table 4 shows that CDPNet outperforms 416 the other models on the WL-Disease dataset with statistical significance. Compared to DenseNet161, the baseline model, CDPNet achieves an accuracy of 95.83%, representing improvements of 2.35%, 3.02%, and 3.65% over Deformer ProtoPNet, ProtoPNet, and DenseNet161, respectively. Figure 7 shows that 419 throughout the entire training process, the CDPNet model consistently outperforms the other four models in both accuracy and loss values, further validating its faster convergence speed. In Figure 8, we explore the effect of varying the prototype count per class on classification performance. CDPNet achieves optimal classification accuracy (95.88%) with 2×2 prototypes configuration, outperforming models with other prototype settings. Therefore, 2×2 prototypes was adopted for all subsequent experiments. Figure 9 presents the sensitivity analysis

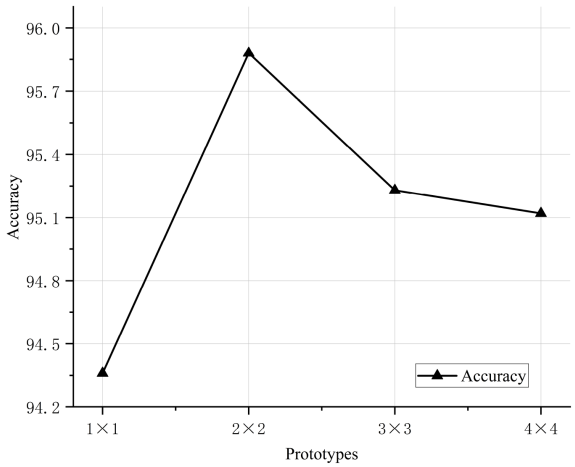

FIGURE 8
Impact of CDPNet-DenseNet161 to the number of prototypes selected on accuracy.

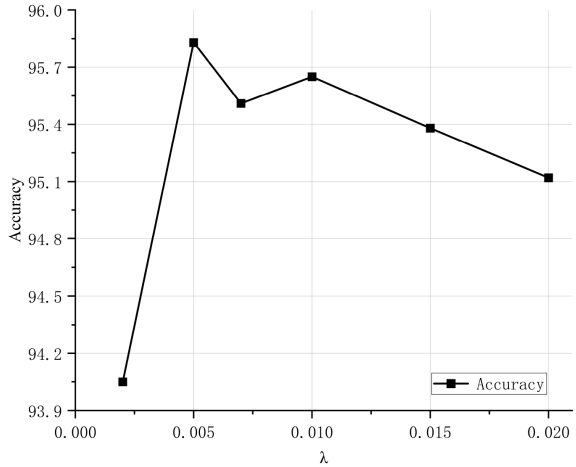

FIGURE 9
Sensitivity of the Barlow Twins component in CDPNet-DenseNet161 to hyperparameter λ.

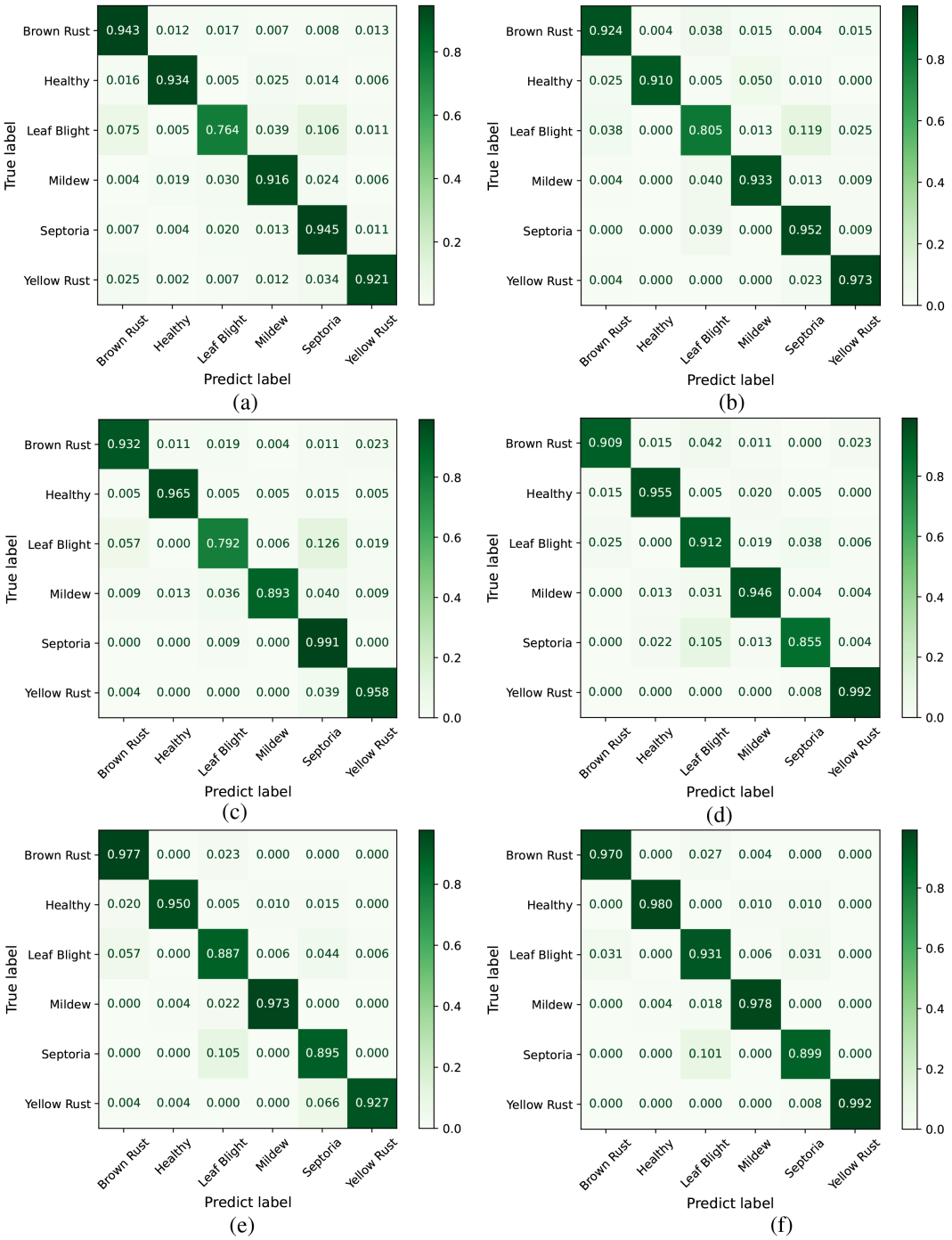

FIGURE 10
Comparison of confusion matrices across different models. **(a)** VGG19, **(b)** ResNet152, **(c)** DenseNet161, **(d)** ProtoPNet-DenseNet161, **(e)** Deformer ProtoPNet-DenseNet161, **(f)** CDPNetDenseNet161.

of CDPNet, based on the DenseNet161 backbone, with respect to its key components. Figure 9 demonstrates the sensitivity of the Barlow Twins component to the hyperparameter λ, which governs the trade-off between invariance and information density in the embedding space. The results indicate that the Barlow Twins are relatively insensitive to this hyperparameter.

Figure 10 shows a confusion matrix that intuitively represents the relationship between predicted results and actual class labels. This illustrates the effectiveness of the model's classification capabilities. In Figure 10a, Leaf Blight exhibits the lowest classification accuracy (76.4%), with 10.6% of test images being misclassified as Septoria and 7.5% misclassified as Brown Rust. In Figure 10d, Septoria has the lowest classification accuracy (85.5%), where 10.5% of test images were incorrectly classified as Leaf Blight. This phenomenon stems primarily from two factors: On one hand, Leaf Blight exhibits a dispersed feature distribution within the

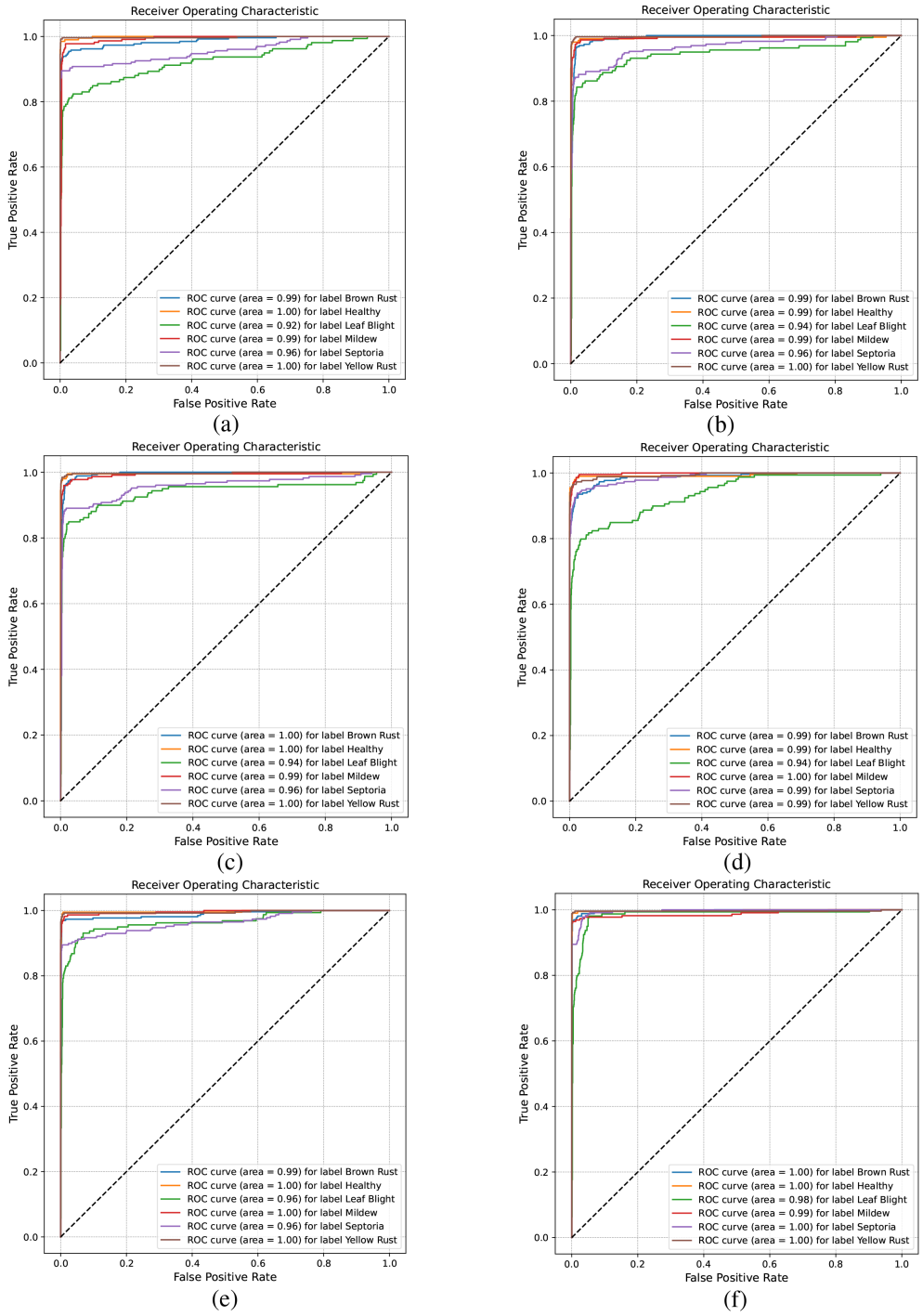

FIGURE 11
Comparison of ROC across different models. **(a)** VGG19, **(b)** ResNet152, **(c)** DenseNet161, **(d)** ProtoPNet-DenseNet161, **(e)** Deformer ProtoPNet-DenseNet161, **(f)** CDPNet-DenseNet161.

dataset, lacking distinct clustered patterns that complicate accurate model recognition. On the other hand, Septoria shares highly similar disease characteristics with Leaf Blight, with significant overlaps in visual features such as morphology and coloration, further exacerbating classification challenges. Compared to other models, the deeper colors along the diagonal of CDPNet's confusion matrix indicate that the majority of classification outcomes are concentrated there. This suggests that the CDPNet model achieves higher recognition accuracy for various diseases, particularly for those with dispersed and easily confused disease regions, such as Leaf Blight and Septoria.

The ROC curve in Figure 11 helps analyze classification performance across different threshold settings. When comparing the ROC curves of different models, those with a higher AUC

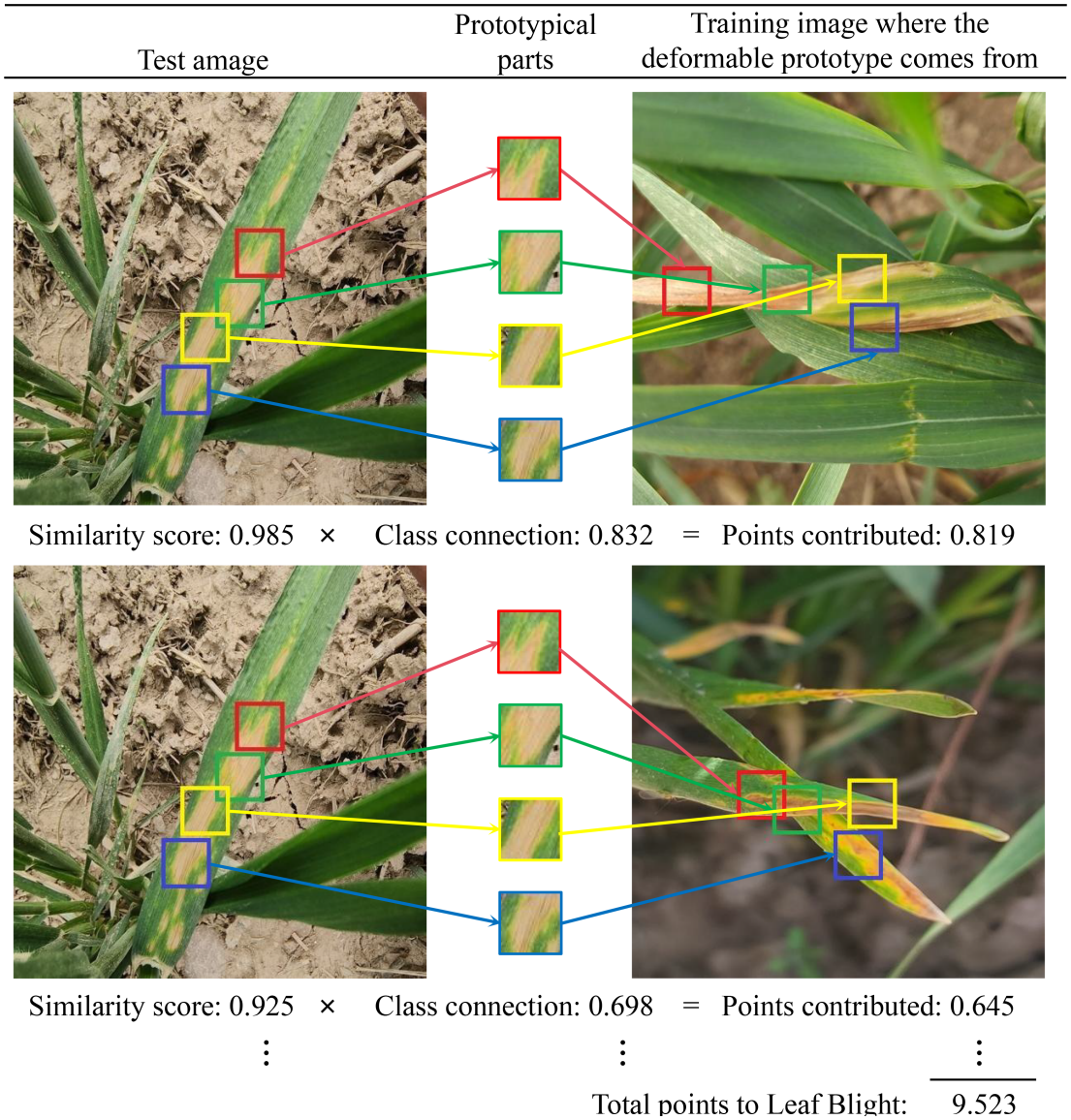

FIGURE 12
The reasoning process of a CDPNet with 2×2 deformable prototypes.

indicate better performance. As shown in Figures 11a–e, the AUC values for Leaf Blight and Septoria leaf diseases are comparatively low. From a phytopathological perspective, Leaf Blight and Septoria diseases are often misidentified in the field. This is primarily due to their highly similar visual symptoms, including leaf necrosis and the yellow halo resulting from chlorophyll degradation, which makes reliable visual differentiation difficult. In contrast, Figure 11f shows that CDPNet demonstrated the highest AUC, achieving superior identification accuracy for these commonly confused diseases. Experimental results indicate that the introduction of the CA mechanism and Barlow twin contrastive learning enabled CDPNet to achieve deeper feature learning for wheat leaf diseases. First, the CA mechanism allows adaptive learning of feature weights across channels, effectively amplifying responses to key disease-related features (e.g., lesion texture, color changes) while suppressing background noise. Second, contrastive learning maximizes similarity between different transformations of the same image while minimizing similarity between different images, thereby optimizing feature relationships across samples and enhancing feature discriminability. As a result, CDPNet improves recognition accuracy for the commonly confused Leaf Blight and Septoria diseases. Moreover, CDPNet's interpretable outputs (Figures 12, 13) help agronomists distinguish these diseases by highlighting specific visual patterns used by the model (e.g., lesion shape, distribution), potentially revealing features that are challenging for the human eye to discern.

4.3.3 K-fold cross-validation

To further validate the model's performance stability on the WL-Disease dataset, we employed k-fold cross-validation,

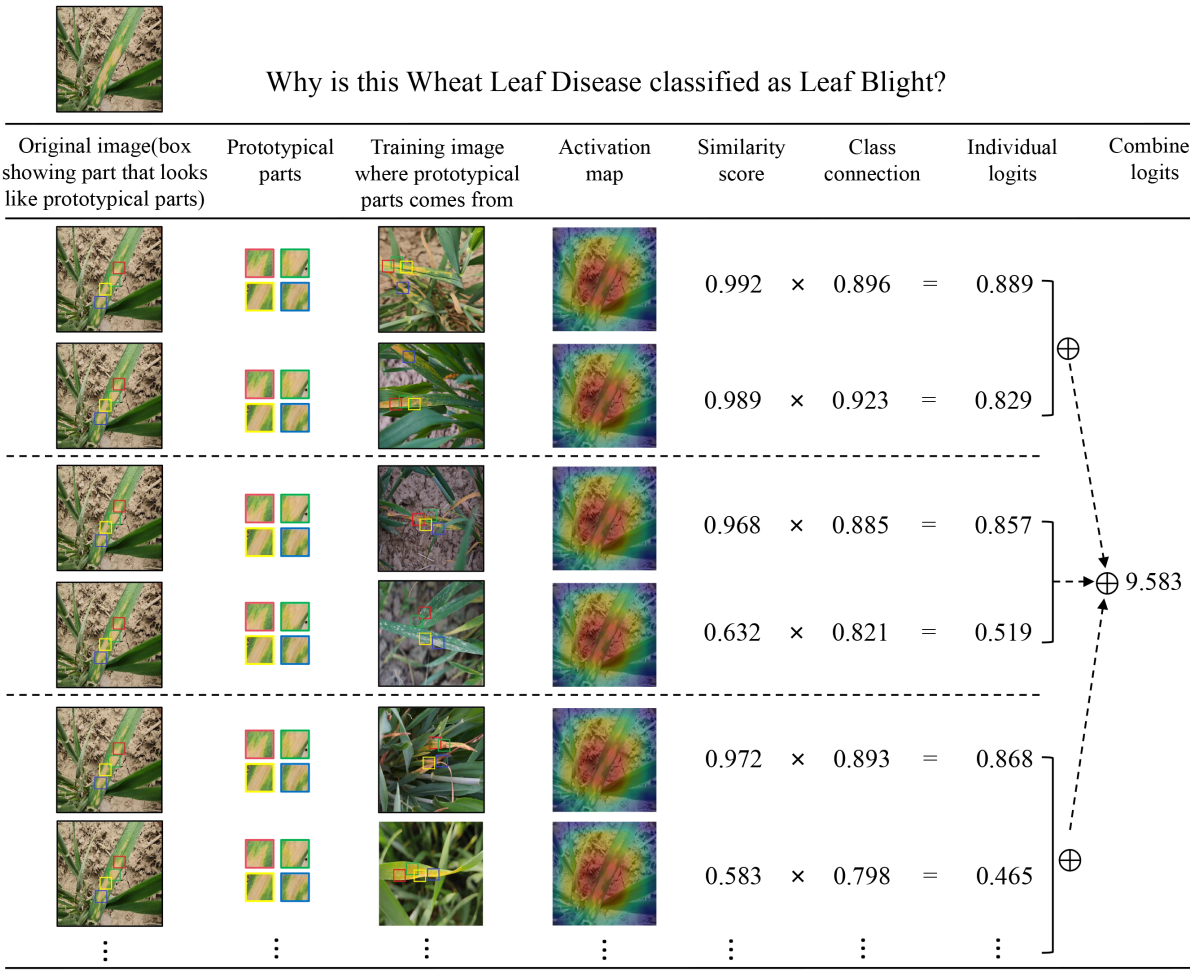

FIGURE 13
The reasoning process of CDPNet in deciding the species of the wheat leaf blight.

processing the dataset sequentially and randomly dividing it into four parts. In each partition, 20% of the data was used as the test set, while the remaining 80% was combined with the other three parts to create a new training set. This approach ensured that each part served as the test set for one partition. We selected DenseNet161 as the baseline model, trained the CDPNet on the training set, validated it on the test set, and recorded the results. Table 5 displays the results of the 5-fold cross-validation. The WL-Disease dataset achieved an average accuracy of 95.83%, with accuracy fluctuations not exceeding 2% across the cross-validation. The results indicate that CDPNet demonstrates stable performance across different subsets, showcasing strong robustness and excellent generalization ability. The model is not prone to significant performance fluctuations due to changes in data partitioning. This suggests that the model does not overfit to specific subsets but learns general features from the data, exhibiting outstanding generalization performance.

TABLE 5 CDPNet+DenseNet161 test results based on k-fold cross-validation.

No of fold	Accuracy (%)
1-flod	95.22
2-flod	95.56
3-flod	96.13
4-fold	95.89
5-fold	96.36
Average	95.83(± 0.61)

4.3.4 Ablation experiments

To further evaluate the effectiveness of the optimization strategies proposed in this study, ablation experiments were performed. The corresponding results are presented in Table 6, which highlights the contribution of each optimization strategy to model performance. Evaluation metrics included accuracy, precision, recall, F1-score on the test set, as well as the number of model parameters. As shown in Table 6, the incorporation of the CA mechanism and the contrastive loss function improved the model's recognition accuracy. Compared with the original Deformer ProtoPNet and using DenseNet161 as the baseline, the CDPNet model, integrating both the CA mechanism

TABLE 6 CDPNet results of ablation experiment.

Model	Cross Attention	Barlow twin loss	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC (%)
Deformer ProtoPNet+VGG19			92.15	91.83	91.52	91.65	98.12
Deformer ProtoPNe+ResNet152			92.63	92.27	92.11	92.19	98.25
Deformer ProtoPNet+DenseNet161			93.48	93.13	92.89	92.99	98.52
Deformer ProtoPNet+VGG19	√		92.85	92.35	92.58	92.45	98.31
Deformer ProtoPNe+ResNet152	√		93.38	92.72	92.97	90.80	98.46
Deformer ProtoPNet +DenseNet161	√		94.13	93.55	93.96	93.62	98.85
Deformer ProtoPNet+VGG19		√	93.64	93.32	93.07	93.13	98.58
Deformer ProtoPNe+ResNet152		√	94.26	93.66	93.62	93.73	99.25
Deformer ProtoPNet+DenseNet161		√	95.25	94.77	95.06	94.83	99.30
Deformer ProtoPNet+VGG19	√	√	94.22	93.72	93.97	93.77	99.16
Deformer ProtoPNe+ResNet152	√	√	94.89	94.21	94.47	94.29	99.38
Deformer ProtoPNet+DenseNet161	√	√	95.83	95.32	95.07	95.13	99.45

"√" indicates that this module has been added.

and the contrastive loss function, achieved an accuracy of 95.83%, representing an improvement of 2.35%. Furthermore, the precision, recall, F1-score, and AUC improved by 2.22%, 2.18%, 2.14%, and 0.93%, respectively. These findings confirm that the integration of the CA mechanism and the contrastive loss function not only avoided adverse effects but also substantially enhanced the recognition performance of CDPNet.

4.3.5 Experimental comparison of public datasets

To validate the generalization ability of the improved CDPNet model, a series of comparative experiments were performed on the PlantVillage and LWDCD 2020 datasets, alongside our self-built dataset. PlantVillage is an open-source plant disease dataset constructed based on image collection of plant leaves. These images were captured under controlled environmental conditions and cover 14 different species of plant. The dataset comprises approximately 54,305 images, categorized into 38 plant disease classes and 1 background image category. For our model training, we selected image data of three different diseases, such as Apple and Corn diseases, from the PlantVillage dataset. The LWDCD 2020 dataset for wheat diseases consists of nearly 7,000 relatively distinct close-up images of wheat diseases, categorized into 12 classes of common wheat diseases in China based on different disease types. Given that our task is wheat leaf disease identification, we selected five kinds of such diseases for model training. Using DenseNet161 as the baseline model, we trained the CDPNet on the training sets of the three datasets and validated it on the corresponding test sets, recording the validation results. Table 7 presents the experimental results of the CDPNet model on the three datasets.

4.3.6 CDPNet interpretability analysis

As an interpretable model, CPDNet not only predicts leaf disease categories but also identifies key affected regions that influence model decisions, enabling explainable image classification and recognition of wheat leaf diseases. Figure 12 illustrates how CPDNet identifies evidence of leaf blight in the test image by comparing its latent features with each variable prototype within the category (each prototypical part is displayed in the "Prototypical parts" column). As shown in Figure 13, when variable prototypes scan the input image, they adaptively adjust their spatial positions. Then, the Prototype similarity scores are computed for each center position using Equation 6. Subsequently,

TABLE 7 CDPNet performance on public datasets.

Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC (%)
PlantVillage-3	92.55	92.69	92.95	92.68	97.83
LWDCD 2020-5	93.35	92.89	93.61	92.78	98.15
WL-Disease	95.83	95.32	95.07	95.13	99.45

the maximum score across all spatial positions is selected using Equation 7 to generate a single "similarity score" for the prototype. This similarity score is multiplied by the class connection score from the fully connected layer to yield the prototype's contribution to the classification result. Finally, the contribution scores of all prototypes are summed to obtain the final classification score for the category. Figures 12, 13 clearly demonstrate that CPDNet can accurately identify regions most affected by Leaf Blight, facilitating the classification and identification of wheat leaves. As a result, CPDNet's interpretable output mechanism offers agronomists an intuitive visualization tool, enabling them to focus on specific visual features (e.g., lesion morphology, spatial distribution) and uncover potential diagnostic characteristics that are challenging to detect through traditional visual inspection.

5 Conclusion

This work introduces a novel deep learning model with intrinsic interpretability for the identification of wheat leaf diseases. Specifically, we present the CDPNet approach, which identifies key regions influencing model decisions by calculating similarity values between convolutional feature maps and latent prototype feature representations. CDPNet incorporates a CA mechanism to effectively isolate target diseased regions from complex backgrounds, thereby enhancing the model's feature extraction capabilities. To address the limited availability of wheat leaf disease image data, we employ a self-supervised contrastive learning approach to capture cross-sample features, thereby improving model efficiency. To validate the model's effectiveness, systematic experiments were conducted using both our self-constructed WL-Disease dataset and two public datasets. The results demonstrate that the proposed CDPNet not only achieves significantly higher accuracy than baseline methods but also provides an interpretable decision-making bases, offering reliable support for practical wheat disease diagnosis in field settings. In summary, the proposed CDPNet model achieves an average accuracy exceeding 92.55% across all three datasets, showcasing its ability to effectively classify and identify diverse crop diseases in real agricultural scenarios.

Future research will focus on developing pre-trained neural network model weights for large-scale plant pest and disease datasets in real-world agricultural settings. This will facilitate the faster convergence of other models when replacing feature extraction network backbones. This research can further alleviate challenges in pest and disease identification within smart agriculture, promoting the intelligent transformation of agricultural practices.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

JZ: Conceptualization, Data curation, Methodology, Project administration, Validation, Writing – original draft. BJ: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. CS: Funding acquisition, Resources, Supervision, Writing – review & editing. HG: Funding acquisition, Resources, Writing – review & editing. LS: Funding acquisition, Resources, Writing – review & editing. BK: Data curation, Methodology, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was funded by the Anhui Science and Technology University Science Foundation (Grant No. WDRC202103, XWYJ202301), the Key Project of Natural Science Research of Universities in Anhui (Grant No. 2022AH051642), the Research and Development Fund Project of Anhui Science and Technology University (Grant No. FZ230122), the key Discipline Construction Project of Anhui Science and Technology University (Grant No. XK-XJGY002), the Anhui Provincial Department of Education Natural Science Major Project (Grant No. 2023AH040276). The authors gratefully acknowledge the Anhui Science and Technology University Science Foundation (Grant No. WDRC202103, XWYJ202301).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

Alaniz, S., Marcos, D., Schiele, B., and Akata, Z. (2021). "Learning decision trees recurrently through communication," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Los Alamitos, CA, USA. 13518–13527.

Alshammari, H., Gasmi, K., Ben Ltaifa, I., Krichen, M., Ben Ammar, L., and Mahmood, M. A. (2022). Olive disease classification based on vision transformer and cnn models. *Comput. Intell. Neurosci.* 2022, 3998193. doi: 10.1155/2022/3998193

Balakrishna, K., and Rao, M. (2019). Tomato plant leaves disease classification using knn and pnn. *Int. J. Comput. Vision Image Process. (IJCVIP)* 9, 51–63. doi: 10.4018/IJCVIP.2019010104

Bao, W., Fan, T., Hu, G., Liang, D., and Li, H. (2022). Detection and identification of tea leaf diseases based on ax-retinanet. *Sci. Rep.* 12, 2183. doi: 10.1038/s41598-022-06181-z

Bao, W., Yang, X., Liang, D., Hu, G., and Yang, X. (2021a). Lightweight convolutional neural network model for field wheat ear disease identification. *Comput. Electron. Agric.* 189, 106367. doi: 10.1016/j.compag.2021.106367

Bao, W., Zhao, J., Hu, G., Zhang, D., Huang, L., and Liang, D. (2021b). Identification of wheat leaf diseases and their severity based on elliptical-maximum margin criterion metric learning. *Sustain. Comput.: Inf. Syst.* 30, 100526. doi: 10.1016/j.suscom.2021.100526

Borhani, Y., Khoramdel, J., and Najafi, E. (2022). A deep learning based approach for automated plant disease classification using vision transformer. *Sci. Rep.* 12, 11554. doi: 10.1038/571s41598-022-15163-0

Chakrabarty, A., Ahmed, S. T., Islam, M. F. U., Aziz, S. M., and Maidin, S. S. (2024). An interpretable fusion model integrating lightweight cnn and transformer architectures for rice leaf disease identification. *Ecol. Inf.* 82, 102718. doi: 10.1016/j.ecoinf.2024.102718

Chang, D. (2025). Vocal performance evaluation of the intelligent note recognition method based on deep learning. *Sci. Rep.* 15, 13927. doi: 10.1038/s41598-025-99357-2

Chang, S., Yang, G., Cheng, J., Feng, Z., Fan, Z., Ma, X., et al. (2024). Recognition of wheat rusts in a field environment based on improved densenet. *Biosyst. Eng.* 238, 10–21. doi: 10.1016/j.biosystemseng.2023.12.016

Chen, C.-F. R., Fan, Q., and Panda, R. (2021a). "Crossvit: Cross-attention multi-scale vision transformer for image classification," in *Proceedings of the IEEE/CVF international conference on computer vision*, Los Alamitos, CA, USA. 357–366.

Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., et al. (2021b). "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 12299–12310.

Dai, G., Tian, Z., Fan, J., Sunil, C., and Dewi, C. (2024). Dfn-psan: Multi-level deep information feature fusion extraction network for interpretable plant disease classification. *Comput. Electron. Agric.* 216, 108481. doi: 10.1016/j.compag.2023.108481

Deng, H., Chen, Y., and Xu, Y. (2025). Ald-yolo: A lightweight attention detection model for apple leaf diseases. *Front. Plant Sci.* 16. doi: 10.3389/fpls.2025.1616224

Dong, T., Ma, X., Huang, B., Zhong, W., Han, Q., Wu, Q., et al. (2024). Wheat disease recognition method based on the sc-convnext network model. *Sci. Rep.* 14, 32040. doi: 10.1038/s41598-024-83636-5

Donnelly, J., Barnett, A. J., and Chen, C. (2022). "Deformable protopnet: An interpretable image classifier using deformable prototypes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10265–10275.

Gao, W., Xu, F., and Zhou, Z.-H. (2022). Towards convergence rate analysis of random forests for classification. *Artif. Intell.* 313, 103788. doi: 10.1016/j.artint.2022.103788

Goethals, S., Martens, D., and Evgeniou, T. (2022). The non-linear nature of the cost of comprehensibility. *J. Big Data* 9, 30. doi: 10.1186/s40537-022-00579-2

Hassan, A., Mumtaz, R., Mahmood, Z., Fayyaz, M., and Naeem, M. K. (2024). Wheat leaf localization and segmentation for yellow rust disease detection in complex natural backgrounds. *Alexandria Eng. J.* 107, 786–798. doi: 10.1016/j.aej.2024.09.018

Hernández, I., Gutiérrez, S., Barrio, I., Íñiguez, R., and Tardaguila, J. (2024). In-field disease symptom detection and localisation using explainable deep learning: Use case for downy mildew in grapevine. *Comput. Electron. Agric.* 226, 109478. doi: 10.1016/j.compag.2024.109478

Javidan, S. M., Banakar, A., Vakilian, K. A., and Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic k-means clustering and machine learning. *Smart Agric. Technol.* 3, 100081. doi: 10.1016/j.atech.2022.100081

Jia, L., Wang, T., Chen, Y., Zang, Y., Li, X., Shi, H., et al. (2023). Mobilenet-ca-yolo: An improved yolov7 based on the mobilenetv3 and attention mechanism for rice pests and diseases detection. *Agriculture* 13, 1285. doi: 10.3390/agriculture13071285

Jiang, Y., Mehta, D., Feng, W., and Ge, Z. (2025). Enhancing interpretable image classification through llm agents and conditional concept bottleneck models. *arXiv preprint arXiv:2506.01334*. doi: 10.48550/arXiv.2506.01334

Jiang, Z., Dong, Z., Jiang, W., and Yang, Y. (2021). Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning. *Comput. Electron. Agric.* 186, 106184. doi: 10.1016/j.compag.2021.106184

Khan, A., Rauf, Z., Sohail, A., Khan, A. R., Asif, H., Asif, A., et al. (2023a). A survey of the vision transformers and their cnn-transformer based variants. *Artif. Intell. Rev.* 56, 2917–2970. doi: 10.1007/s10462-023-10595-0

Khan, F., Zafar, N., Tahir, M. N., Aqib, M., Waheed, H., and Haroon, Z. (2023b). A mobile-based system for maize plant leaf disease detection and classification using deep learning. *Front. Plant Sci.* 14. doi: 10.3389/fpls.2023.1079366

Khan, H., Haq, I. U., Munsif, M., Mustaqeem,, Khan, S. U., and Lee, M. Y. (2022a). Automated wheat diseases classification framework using advanced machine learning technique. *Agriculture* 12, 1226. doi: 10.3390/agriculture12081226

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022b). Transformers in vision: A survey. *ACM Comput. Surveys (CSUR)* 54, 1–41. doi: 10.1145/3505244

Kim, M., Kim, H.-I., and Ro, Y. M. (2024). Prompt tuning of deep neural networks for speaker-adaptive visual speech recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* pp. 1042–1055. doi: 10.1109/TPAMI.2024.3484658

Lin, H., Cheng, X., Wu, X., and Shen, D. (2022). "Cat: Cross attention in vision transformer," in *2022 IEEE international conference on multimedia and expo (ICME) (IEEE)*. Piscataway, NJ, USA. 1–6.

Ma, C., Donnelly, J., Liu, W., Vosoughi, S., Rudin, C., and Chen, C. (2024). "Interpretable image classification with adaptive prototype-based vision transformers," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*. Red Hook, NY, USA. pp. 41447–41493.

Marcus, E., and Teuwen, J. (2024). Artificial intelligence and explanation: How, why, and when to explain black boxes. *Eur. J. Radiol.* 173, 111393. doi: 10.1016/j.ejrad.2024.111393

Mishra, R., Rajpal, A., Bhatia, V., Rajpal, S., Agarwal, M., et al. (2024). I-ldd: an interpretable leaf disease detector. *Soft Comput.* 28, 2517–2533. doi: 10.1007/s00500-023-08512-2

Nawaz, M., Nazir, T., Javed, A., Amin, S. T., Jeribi, F., and Tahir, A. (2024). Coffeenet: A deep learning approach for coffee plant leaves diseases recognition. *Expert Syst. Appl.* 237, 121481. doi: 10.1016/j.eswa.2023.121481

Nigam, S., Jain, R., Marwaha, S., Arora, A., Haque, M. A., Dheeraj, A., et al. (2023). Deep transfer learning model for disease identification in wheat crop. *Ecol. Inf.* 75, 102068. doi: 10.1016/j.ecoinf.2023.102068

Pattnaik, G., and Parvathi, K. (2021). "Automatic detection and classification of tomato pests using support vector machine based on hog and lbp feature extraction technique," in *Progress in Advanced Computing and Intelligent Engineering: Proceedings of ICACIE 2019*, Springer, Singapore. 2, 49–55. doi: 10.1007/978-981-15-6353-95

Quan, W., Zhang, R., Zhang, Y., Li, Z., Wang, J., and Yan, D.-M. (2022). Image inpainting with local and global refinement. *IEEE Trans. Image Process.* 31, 2405–2420. doi: 10.1109/TIP.2022.3152624

Raval, H., and Chaki, J. (2024). Ensemble transfer learning meets explainable ai: A deep learning approach for leaf disease detection. *Ecol. Inf.* 84, 102925. doi: 10.1016/j.ecoinf.2024.102925

Rezvani, S., and Wu, J. (2023). Handling multi-class problem by intuitionistic fuzzy twin support vector machines based on relative density information. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 14653–14664. doi: 10.1109/TPAMI.2023.3310908

Simón, M. R., Börner, A., and Struik, P. C. (2021). Fungal wheat diseases: etiology, breeding, and integrated management. *Frontiers in plant science* 12, 671060. doi: 10.3389/fpls.2021.671060

Syazwani, R. W. N., Asraf, H. M., Amin, M. M. S., and Dalila, K. N. (2022). Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alexandria Eng. J.* 61, 1265–1276. doi: 10.1016/j.aej.2021.06.053

Thakur, P. S., Chaturvedi, S., Khanna, P., Sheorey, T., and Ojha, A. (2023). Vision transformer meets convolutional neural network for plant disease classification. *Ecol. Inf.* 77, 102245. doi: 10.1016/j.ecoinf.2023.102245

Thakur, P. S., Khanna, P., Sheorey, T., and Ojha, A. (2022). Trends in vision-based machine learning techniques for plant disease identification: A systematic review. *Expert Syst. Appl.* 208, 118117. doi: 10.1016/j.eswa.2022.118117

Wani, J. A., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S., and Singh, S. (2022). Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges. *Arch. Comput. Methods Eng.* 29, 641–677. doi: 10.1007/s11831-021-09588-5

Wei, K., Chen, B., Zhang, J., Fan, S., Wu, K., Liu, G., et al. (2022). Explainable deep learning study for leaf disease classification. *Agronomy* 12, 1035. doi: 10.3390/agronomy12051035

Xu, S., Chen, S., Xu, R., Wang, C., Lu, P., and Guo, L. (2024). Local feature matching using deep learning: A survey. *Inf. Fusion* 107, 102344. doi: 10.1016/j.inffus.2024.102344

Xu, Y., Du, B., and Zhang, L. (2021). Self-attention context network: Addressing the threat of adversarial attacks for hyperspectral image classification. *IEEE Trans. Image Process.* 30, 8671–8685. doi: 10.1109/TIP.2021.3118977

Zbontar, J., Jing, L., Misra, I., LeCun, Y., and Deny, S. (2021). "Barlow twins: Self-supervised learning via redundancy reduction," in *International conference on machine learning (PMLR)*, Brookline, MA, USA. 12310–12320.

Zeng, Z., and Xiong, D. (2022). Unsupervised and few-shot parsing from pretrained language models. *Artif. Intell.* 305, 103665. doi: 10.1016/j.artint.2022.103665

Zhang, F., Bao, R., Yan, B., Wang, M., Zhang, Y., and Fu, S. (2024). Lsannet: A lightweight convolutional neural network for maize leaf disease identification. *Biosyst. Eng.* 248, 97–107. doi: 10.1016/j.biosystemseng.2024.09.023

Zhang, L., Ding, G., Li, C., and Li, D. (2023). Dcf-yolov8: an improved algorithm for aggregating low-level features to detect agricultural pests and diseases. *Agronomy* 13, 2012. doi: 10.3390/agronomy13082012

Zhang, Y., Zheng, Y., Wang, D., Gu, X., Zyphur, M. J., Xiao, L., et al. (2025). Shedding light on the black box: Integrating prediction models and explainability using explainable machine learning. *Organizational Res. Methods*, 10944281251323248. doi: 10.1177/10944281251323248

Zhao, Y., Li, Y., Wu, N., and Xu, X. (2024). Neural network based on convolution and self-attention fusion mechanism for plant leaves disease recognition. *Crop Prot.* 180, 106637. doi: 10.1016/j.cropro.2024.106637