Journal of Electronic Research and Application, 2026, Volume 10, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

Research on Crop Leaf Disease Identification and Severity Assessment Based on Lightweight Multitask Deep Networks

Jiali Du*

Business School, Xi'an International Studies University, Xi'an, Shaanxi, China

***Corresponding author:** Jiali Du, 3204243702@qq.com

Abstract: In response to model 1, we first cleaned and standardized 61 crop disease image categories by removing duplicates through comparing image filenames with label files using provided path information. Valid samples were resized and augmented to construct a multi-disease classification model based on the lightweight MobileNetV3-Large, with category IDs mapped to disease names. The model was trained and validated with cross-entropy loss, AdamW optimizer, and cosine annealing learning rate, with epoch-dependent loss and accuracy curves recorded. For model 2, a few-shot recognition solution was developed based on model 1, retaining 10 training samples per category. Using pre-trained MobileNetV3-Large as the feature backbone (parameters < 20M), only upper convolutional and classification layers were fine-tuned. Enhanced augmentation, label smoothing, and cosine annealing mitigated overfitting and class imbalance, achieving ~73% validation accuracy for 61 categories; Grad-CAM confirmed the model focuses on leaf lesions. Regarding model 3, severity-graded prediction was implemented by mapping 61 diseases to 3 severity levels via appendix JSON annotations and disease description tables. Images were regrouped to build a three-classification dataset, and a severity prediction model with MobileNetV3-Large (transfer learning, augmentation) was trained, outputting overall accuracy, macro-F1, recall, and a confusion matrix; Grad-CAM visualized key lesions for high-confidence correct predictions. For model 4, a lightweight integrated multi-task model was developed for simultaneous disease identification and severity assessment, using MobileNetV3-Large as the shared feature backbone with 61-category disease and 3-category severity classification heads. Joint optimization via multi-task loss enabled feature sharing and fine-grained assessment, with joint accuracy, confusion matrices, and Grad-CAM analyzing synergy and lesion focus, supporting interpretable diagnosis reports.

Keywords: MobileNetV3-large convolutional neural network; Transfer learning, Cross-entropy loss function; Multitask learning

1. Introduction

Agriculture is the foundation of human survival, providing food and other essential necessities, and serves as the fundamental guarantee for a nation's economy. However, traditional agriculture has long faced persistent challenges such as reduced yields, diminished quality, and economic losses caused by crop disease outbreaks. With continuous advancements in technology, the advent of the Smart Agriculture 4.0 era has brought revolutionary changes to the agricultural sector. The application of technologies like drone remote sensing and IoT devices has significantly enhanced farm management efficiency, largely resolving longstanding issues in traditional agriculture, such as yield and quality declines, as well as economic losses from pest and disease outbreaks. Although China has achieved remarkable progress in agricultural automation, the current intelligent upgrade mechanisms still require continuous refinement to overcome the limitation of prioritizing automation over predictability.Therefore, to achieve a leap from reactive remediation to precise prevention and control through accurate prediction of crop growth trends, this paper focuses on constructing an intelligent crop health management system leveraging artificial intelligence and computer vision technologies.

2. Model development and solution

2.1. Model establishment

To achieve precise classification of a dataset containing images of 61 types of agricultural diseases, the data underwent cleaning, augmentation, and normalization according to the task requirements.

2.1.1. Input preprocessing module

Input images are first uniformly resized to 224×224 for normalization. A 3×3 convolutional kernel with stride 2 and padding 1 is applied to achieve spatial downsampling. Feature data is then batch-standardized to effectively suppress distribution shifts and accelerate model convergence.

2.1.2. Feature extraction module

This module comprises multiple groups of depthwise separable convolutions and pointwise convolutions. The 1×1 convolutio ns handle channel expansion, while the 3×3 convolutions perform feature extraction. Additionally, an SE attention module is incorporated alongside a hard-swish activation function to enhance feature extraction capabilities while improving computational efficiency and quantization friendliness [1]. This approach prevents information loss during feature compression and enhances sensitivity to key disease image features through dynamic weight allocation. The specific calculation formula is shown in **Equation (1)**:

$$\text{hard-swish}[y_i]=y_i\frac{\min(\max\left(0,y_i+3\right),6)}{6} \tag{1}$$

Here, y^i represents one input value to the activation function.

2.1.3. Disease image recognition module

This module consists of an average pooling layer and two fully connected layers. First, average pooling compresses the feature channels to extract high-level semantic features. Second, based on the trained MobileNetV3 model, visualization methods like Grad-CAM were applied to generate attention heatmaps for representative leaf images. Finally, to further validate the accuracy of the constructed model for crop disease image classification, this paper employs two evaluation metrics, accuracy and loss value, to assess model reliability. Accuracy measures the model's predictive precision for positive cases, reflecting the reliability of crop disease identification. The specific calculation formulas are shown in **Equations (2)** and **(3)**:

$$\text{accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{2}$$

$$\text{loss}=-\sum_{i=1}^{n}\hat{y}_{i1}\log y_{i1}+y_{i2}\log y_{i2}+\ldots+\hat{y}_{in}\log y_{in} \tag{3}$$

Among these, accuracy represents the model's accuracy rate, loss represents the loss function, precision represents the model's precision rate, and recall represents the model's recall rate.

2.2. Model solving

2.2.1. Solving crop disease image classification based on the MobileNetV3-Large convolutional neural network model

Based on the constructed model, Python was used to input the processed and augmented image data into the model for training and classification. This achieved high-accuracy classification for a dataset of 61 categories (actually 60 categories) of agricultural disease images. Attention heatmaps were generated for several representative leaf images.

As shown in **Figure 1**, the deeper the red color of the hotspots indicates a stronger focus intensity. The red areas in the image are predominantly concentrated on the diseased spots of the leaves.

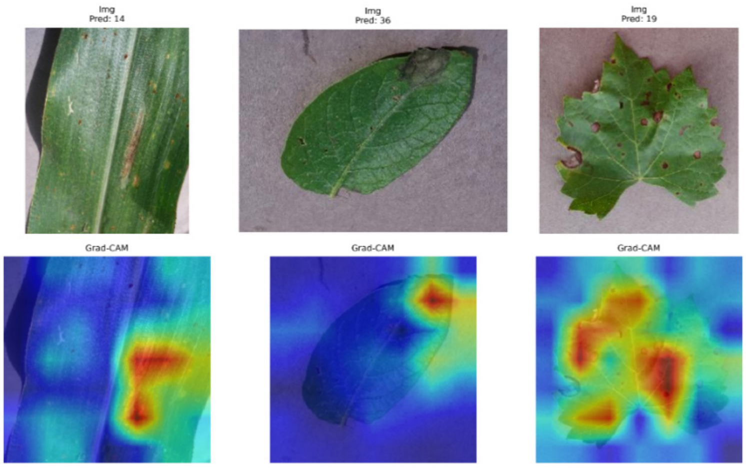

Figure 1. Representative leaf image attention heatmap.

2.2.2. Evaluate the classification accuracy of the model using two metrics: accuracy and loss value

Based on the classification results output by the model, substitute the data obtained from the training set into **Equations (3)** and **(4)** to calculate the classification accuracy and loss value. Simultaneously, calculate the accuracy and loss value for the validation data using the validation set provided in the appendix. The accuracy curves and loss value curves for the training set and validation set are shown in **Figure 2**. As indicated by the curve trends, their accuracy and loss values exhibit numerical discrepancies due to the significant difference in data volume between the two datasets. However, the accuracy and loss values of the training and validation sets largely follow the same trend: both sets show a decreasing trend in loss values at the same level, while their accuracy increases at the same level.

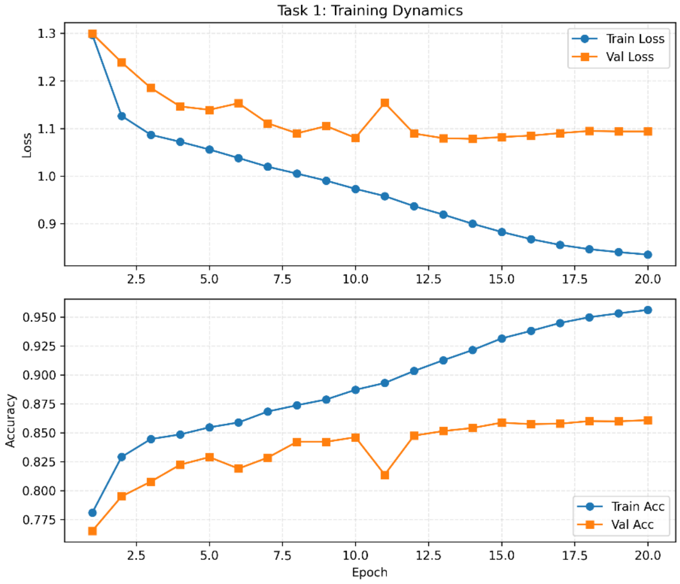

Figure 2. Accuracy and loss value curve.

3. Model 2 development and solution

3.1. Model establishment

First, the problem requires randomly selecting 10 images per category from the full dataset to form the training set for effectively classifying 61 types of agricultural diseases. The model must not exceed 20 million parameters and cannot utilize additional training data. Compared to model 1, the drastically reduced training set size makes it challenging to ensure classification accuracy. Therefore, this paper employs transfer learning, leveraging pre-trained ImageNet weights and the MobileNetV3-Large architecture validated in model 1. The specific steps are as follows:

(1) Input Preprocessing Module: According to the task requirements, 61 types of agricultural diseases, each with 10 images, were selected from the data in model 1 to form a new training set, which underwent unified processing;

(2) Feature Extraction Module: Building upon the features extracted in model 1, this task employs transfer learning for precise classification of the small-sample crop disease image training set;

(3) Disease Image Recognition and Classification Module: Based on the image data processing and feature value extraction from the previous two steps, disease classification is performed on the training set according to the extracted feature values. The training process of the transfer learning-based crop disease image classification model is illustrated in **Figure 3**.

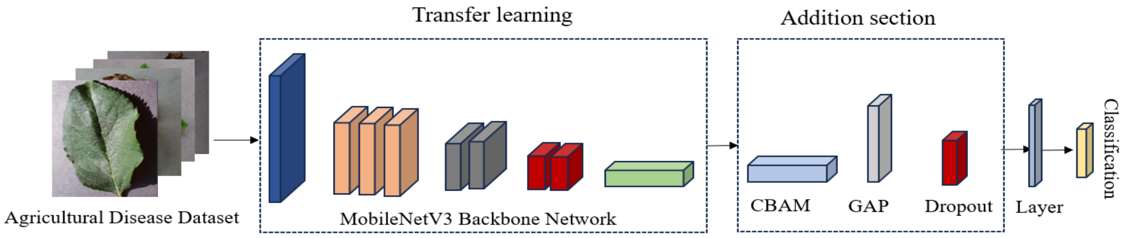

Figure 3. Model construction flowchart.

Second, based on the trained MobileNetV3 model using transfer learning, we continue to employ the Grad-CAM visualization method to generate attention heatmaps for several representative leaf images, as shown in **Figure 3**. This allowed us to observe whether the network truly focused on lesion areas rather than background noise, thereby further validating the effectiveness of the constructed model for data classification at the interpretability level.

Finally, to further validate the accuracy of the model constructed for crop disease image classification as addressed in model 2, we continued to use two evaluation metrics, accuracy and loss value, to assess the model's reliability.

3.2. Model solving

Solving crop disease image classification using the MobileNetV3-Large convolutional neural network model under a few-shot transfer learning approach. Given the limited sample size in this task, transfer learning is employed to address the challenge of declining classification accuracy in few-shot scenarios. As this task also involves classifying crop disease images, the network model constructed in model 1 is further adapted to solve model 2 based on the transfer learning methodology. Using Python, the processed and augmented image data were fed into the constructed model for training and classification. The solution process largely eliminated interference from other factors and met the requirements for training set size and time constraints. It achieved high-accuracy classification for a dataset comprising 61 categories (actually 60 categories), each containing only 10 agricultural disease images.

The heatmap reveals a high degree of spatial correspondence between the red response areas and the actual leaf disease occurrence zones (**Figure 4**). As the severity of leaf disease increases, the color gradient in the corresponding heatmap regions deepens significantly, demonstrating a clear positive correlation. This further indicates that under small-sample dataset conditions, the MobileNetV3-Large convolutional neural network model optimized using transfer learning strategies maintains excellent classification accuracy for leaf disease images.

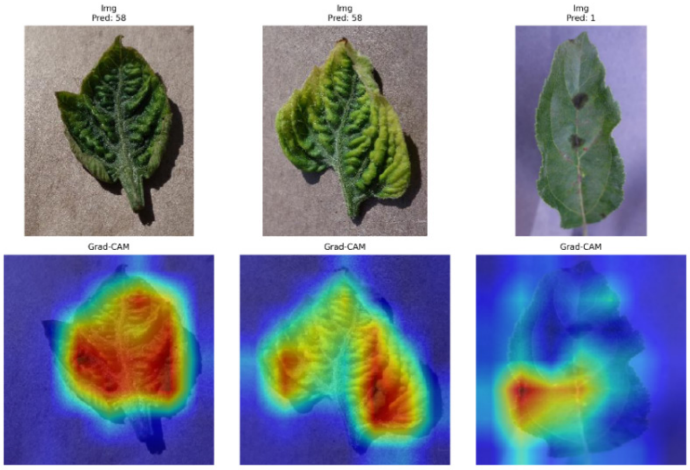

Figure 4. Representative leaf image attention heatmap.

As shown in **Figure 5**, there exists a significant disparity in the absolute values of accuracy and loss between them due to the difference in sample sizes between the two datasets. However, from the perspective of dynamic trends, the accuracy curves and loss curves of the training and validation sets consistently maintain a high degree of alignment.

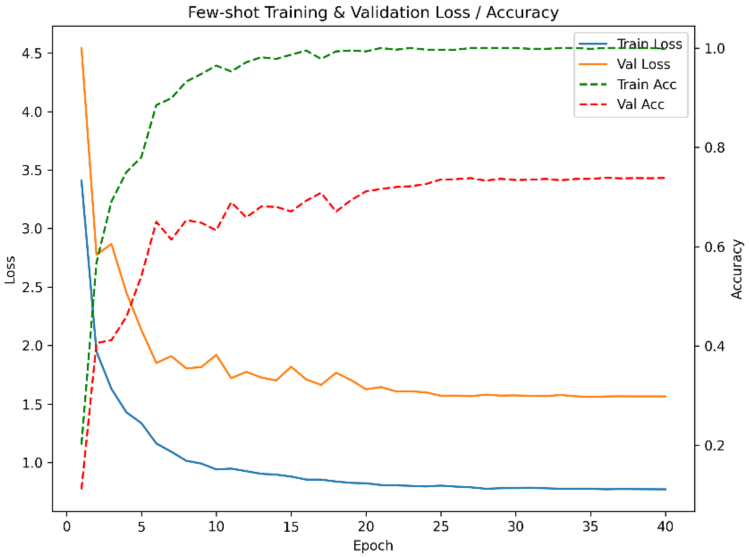

Figure 5. Accuracy and loss value curve.

4. Model 3 development and solution

4.1. Model establishment

To classify the severity of diseases across 10 crops into three categories, we identified the disease_class and its corresponding Chinese name based on the labels in the attached JSON file as specified in the task requirements. We then automatically mapped the original 61 disease categories to three severity levels: those with “healthy” in their names were mapped to “healthy,” those with ‘severe’ in their names were mapped to "severe disease," and all other diseases without explicitly labeled severity were uniformly classified as "mild disease." Additionally, considering that disease severity classification is a fine-grained visual recognition task requiring training within limited time constraints. When facing insufficient data, employing transfer learning not only reduces computational resources required for model training but also enhances the generalization capability of the recognition model [2]. The specific model construction is as follows:

(1) Autonomous mapping of disease severity data into three categories: Based on preprocessed and augmented image data, combined with the mapping rules for the "disease_class" tag and its corresponding Chinese names from the provided JSON file, an initial three-class classification of crop disease severity is completed, categorizing image data into healthy, mild disease, and severe disease. Using this three-class image data as the foundation, new training and validation sets are constructed;

(2) Input Preprocessing Module: First, obtain a pre-trained model for three-class classification from the new dataset after the initial classification. Then, perform weight transfer training, adding a new classifier for identifying disease severity categories. Train this classifier using the augmented leaf disease dataset to obtain the disease severity recognition model. Since the experimental model is a lightweight convolutional neural network with fewer parameters, no layers are frozen during training. Instead, the weight parameters of the pre-trained network model are trained [3];

(3) Feature extraction: Input images undergo standard convolution and are mapped to a high-dimensional feature space via three classification channels;

Finally, to further validate the accuracy of the model constructed for crop disease severity classification, this study calculates and plots accuracy and loss curves. This measures the model's predictive precision for positive cases, reflecting the reliability of crop disease severity classification. Simultaneously, recall curves were calculated and plotted for each category to assess recognition completeness within specific severity levels.

$$\text{precision}=\frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

$$\text{recall}=\frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

Among these, precision represents the accuracy of the model's classification of disease severity levels, while recall represents the completeness of classification for a specific category.

4.2. Model solving

A three-class classification model for crop disease severity was developed using the MobileNetV3-Large convolutional neural network and transfer learning. Core operations primarily consist of 3×3 depth convolutions and 1×1 pointwise convolutions, with an initial kernel count of 16. The activation function employs h-swish, and stride is set to 2. Image features are mapped to a high-dimensional feature space through three classification channels. Following feature mapping, the data undergoes global average pooling, compressing the 7×7×960 feature map into a 1×1×960 vector. A 1×1 convolution then adjusts the channel dimension to 1280, with the final three-class classification results output via a fully connected layer.

The Grad-CAM heatmap visualization reveals that healthy leaves exhibit the smallest proportion of high-response red areas, with no distinct concentration patterns (**Figure 6**). Leaves with mild disease show red heatmap regions precisely focused on the lesion sites. Crucially, leaves with severe disease consistently display the highest heatmap coefficients concentrated in the core infection zones and extensive necrotic areas.

Confusion matrix analysis reveals that in the three-class classification task for crop disease severity, the total number of correctly classified samples across healthy, moderately diseased, and severely diseased categories reached 3,966, with only 567 misclassified samples. This achieves an overall classification accuracy of 87.5%. This demonstrates the constructed model's high classification precision in disease severity grading tasks.

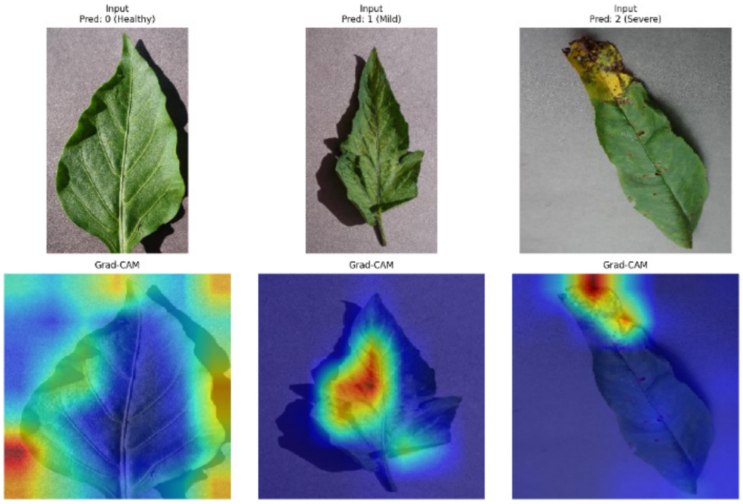

Figure 6. Representative leaf image attention heatmap.

Simultaneously, the validation set data provided in the appendix is used to compute the validation accuracy, loss value, and recall rate. The accuracy curves, loss value curves, and recall rate curves for the training set and validation set are shown in **Figure 7**. The newly constructed validation set achieved an overall classification accuracy of 88.5% and a macro-average F1 score of 87%. Both core metrics are at a high level, indicating the model's excellent comprehensive classification performance in the three-classification task for crop disease severity. It can reliably distinguish between healthy, moderately diseased, and severely diseased samples. Secondly, the cross-entropy loss values for both the training and validation sets exhibit a synchronized downward trend, with the training set loss dropping below 40% in the later stages.

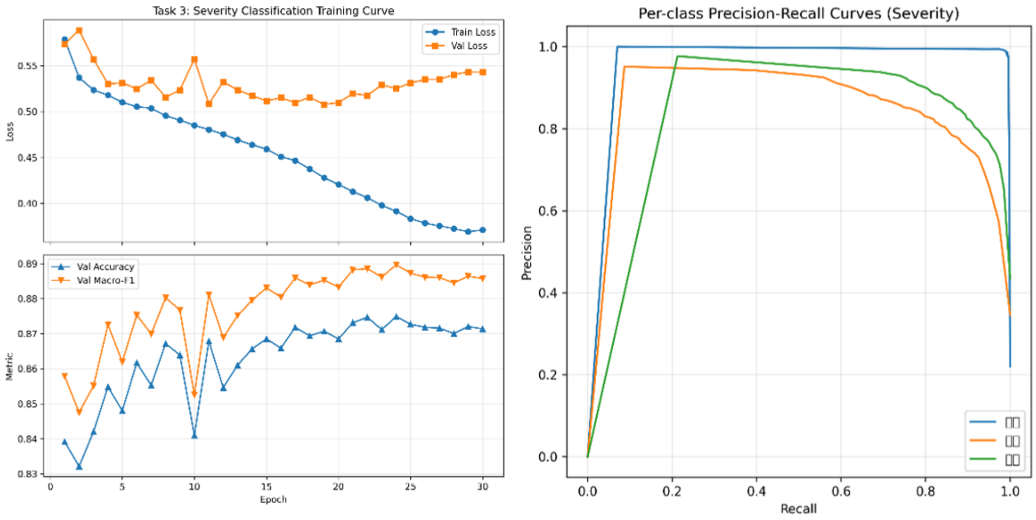

Figure 7. Precision, loss, recall curve.

5. Model 4 development and solution

5.1. Model establishment

To comprehensively classify 61 types of agricultural diseases and assess their severity levels, we utilize the models constructed for models 1, 2, and 3, along with the data in the appendix. In accordance with the requirement for multi-task joint learning, this approach employs the ImageNet-pretrained MobileNetV3-Large as the shared backbone. The specific model architecture is as follows:

(1) MobileNetV3-Large Network Architecture: The input preprocessing module implements dual-label mapping and data augmentation. Based on text descriptions of disease classifications, it automatically constructs a three-level severity mapping rule (healthy, moderate, severe) to achieve consistent conversion from original disease category labels to severity labels;

(2) CBAM Attention Module: Notably, since this task employs a multi-task learning approach, different tasks require the convolutional network to focus on distinct aspects. Therefore, the CBAM attention module is introduced during model construction.

$$Y_c = F_c(Y) \cdot Y \tag{6}$$

$$Y_s = F_s(Y_c) \cdot Y_c \tag{7}$$

Here, Y denotes the input feature map, $F_c(Y)$ denotes the channel attention feature map, Y_c denotes the output feature map, $F_s(Y_c)$ denotes the spatial attention feature map, and F_s denotes the spatial attention output feature map.

(3) Multi-task cross-entropy loss function: Due to differences in sample size, task objectives, and activation functions between the crop disease classification and disease severity classification tasks, Training multi-task convolutional neural networks primarily involves minimizing the total loss function L. T.

$$L = \lambda_{dis} L_{dis} + \lambda_{sev} L_{sev} \tag{8}$$

Here, L represents the total loss function, denotes the loss for the disease classification task,and denotes the loss for the disease severity classification task.

Based on the MobileNetV3-Large convolutional neural network, this study further employs the Grad-CAM visualization method. By constructing Grad-CAM heatmaps based on severity branches, the model's focus regions on leaves are overlaid with gradient information from actual visible lesions to generate transparent heatmaps.

5.2 Model solution

Solved using a multi-task learning model based on the MobileNetV3-Large network. First, based on the aforementioned multi-task learning framework, this approach implements a feature extraction architecture centered on MobileNetV3-Large using Python. It loads pre-trained model weights from the ImageNet-1K dataset, reusing the general visual features learned from large-scale images.

Crucially, leaves with severe disease consistently display maximum heatmap response areas covering the core infection sites and extensive necrotic regions (**Figure 8**). These visualizations conclusively demonstrate that the constructed multi-task classification model can accurately capture species-specific features of different crop diseases and graded characteristics across varying severity levels. It achieves high-precision joint classification for 61 disease categories across 10 crops and 3 severity levels. Simultaneously, the model employs an adaptive attention mechanism to focus decision-making on key disease regions strongly correlated with classification tasks. Only 577 samples exhibited errors in either a single task or both tasks, achieving an overall joint classification accuracy of 87.2%.

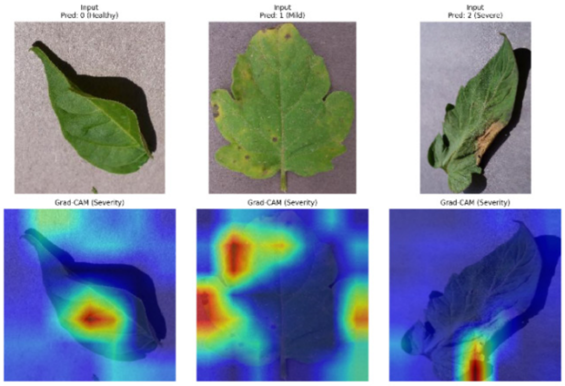

Figure 8. Representative leaf image attention heatmap.

6. Conclusion

This research systematically addresses four core questions related to crop disease diagnosis, developing a series of targeted models based on the lightweight MobileNetV3-Large architecture to achieve accurate identification, few-shot recognition, severity grading, and integrated multi-task diagnosis. The findings demonstrate the feasibility and effectiveness of using lightweight deep learning models for crop disease analysis, providing a practical technical basis for intelligent agricultural diagnosis. For disease identification, the cleaned and standardized dataset combined with data augmentation ensures the model's robustness, while the adopted optimization strategies guarantee stable training performance. In few-shot scenarios, the model maintains ~73% validation accuracy with only 10 samples per category, effectively mitigating overfitting and class imbalance through targeted fine-tuning and augmentation strategies. Grad-CAM visualization consistently confirms that all models focus on key leaf lesion areas, verifying their interpretability. The severity grading model successfully maps 61 disease categories to three severity levels, with comprehensive evaluation metrics and confusion matrices demonstrating its strong discrimination capability. The integrated multi-task model achieves joint optimization of disease identification and severity assessment, leveraging shared feature extraction to enhance efficiency while ensuring fine-grained performance. Overall, this research provides a complete, lightweight, and interpretable technical solution for crop disease diagnosis, laying a solid foundation for generating actionable diagnostic reports and promoting intelligent agricultural development.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Jin Y, Zhen Y, Liu P, 2025, Research on Maize Leaf Disease Identification Based on Transfer-Learned

MobileNetV3. Bulletin of Chinese Agricultural Sciences, 41(23): 145–154.

[2] Ge B, Zhang X, Liu S, et al., 2024, Discrimination of Yellow Tea Fermentation Level Based on the Lightweight Neural Network MobileNetV3–large. Journal of Hunan Agricultural University (Natural Science Edition), 50(1): 91–99.

[3] Wang J, Feng S, Cheng Y, 2021, A Review of Lightweight Neural Network Architectures for Deep Learning. Computer Engineering, 47(8): 1–13.