Academic Education Publishing House

Industry Science and Engineering Vol. 3 No. 5, 2026

Research on Plant Seedling Image Recognition Based on CBAM-EfficientNet-B0

Ke Zhang[1], Ruoxi Cui[2], Xiaofeng Li[3,*]

[1]*College of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou, Henan, China*

[2] *College of Information Science and Engineering, Henan University of Technology, Zhengzhou, Henan, China*

[3]*ANHUI USTC iFLYTEK Co., Ltd, Hefei, Anhui, China*

**Corresponding Author*

Abstract: With the ongoing development of smart agriculture and computer vision, deep learning based approaches for plant recognition have gradually emerged as an important research direction in agricultural intelligence. During the seedling stage, different plant categories often exhibit similar morphological traits and background interference is relatively high. Moreover, traditional convolutional neural networks lack sufficient feature extraction ability. Therefore, we propose a plant seedling image recognition method based on CBAM-EfficientNet-B0. Using EfficientNet-B0 as the backbone network, we integrate a Convolutional Block Attention Module after each MBConv block. By leveraging both channel and spatial attention mechanisms, the proposed model can more effectively concentrate on critical feature regions, thereby enhancing classification performance for plant seedling images. Our experimental results demonstrate that the combination of attention mechanisms with a lightweight convolutional neural network can effectively improve plant seedling image recognition, and also has practical utility for automatic plant identification and classification in smart agriculture.

Keywords: Plant Seedling Recognition; Deep Learning; EfficientNet-B0; CBAM Attention Mechanism; Image Classification

1. Introduction

With the continuous advancement of artificial intelligence and smart agriculture technologies, plant recognition methods based on computer vision have gradually become an important direction in the field of agricultural informatization [1]. In agricultural production, accurate recognition of plants at the seedling stage has a significant impact on crop type management, weed removal, and the automated operation of intelligent agricultural equipment. However, plant seedlings often exhibit similar appearances during early growth, with minor differences in color, texture, and leaf morphology among species. Additionally, factors such as complex backgrounds and illumination variations interfere with recognition results. Traditional manual recognition methods are therefore insufficient to meet the development needs of modern intelligent agriculture [2]. When submitting your final draft to the STEMM Press. These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts.

Early plant image recognition methods mainly relied on manually extracted features such as color, texture, and shape, combined with traditional machine learning algorithms like Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) for classification [3]. Such methods achieved certain results on small-scale datasets, but their feature representation capability was limited, and their adaptability to complex environments was insufficient, leading to problems in practical applications. With the development of deep learning, Convolutional Neural Networks (CNNs), leveraging their powerful automatic feature extraction ability, have achieved remarkable results in image classification, object detection, and medical image analysis [4].

In recent years, lightweight networks have gradually become a research hotspot. Compared with traditional deep convolutional networks, lightweight models can reduce the number of

parameters and computational cost while maintaining recognition accuracy, making them more suitable for deployment on mobile terminals and embedded devices. The EfficientNet model employs a compound scaling strategy to uniformly optimize network depth, width, and input resolution, demonstrating good efficiency and performance in image classification tasks [5]. However, in fine-grained classification tasks of plant seedlings, basic convolutional networks are still easily disturbed by background noise, resulting in insufficient feature extraction from key regions. The emergence of attention mechanisms provides a new solution to this problem. Attention mechanisms can mimic the human visual system by focusing on important regions, allowing the model to pay more attention to target area information. Among them, the Convolutional Block Attention Module (CBAM) combines channel and spatial attention mechanisms, effectively improving the model's feature representation capability. Currently, CBAM has been widely applied in tasks such as image classification, object detection, and medical image analysis, achieving good results.

Therefore, to address the problems of small inter-class differences, strong background interference, and insufficient feature extraction capability of traditional models in plant seedling image recognition, this paper proposes a plant seedling image recognition method based on CBAM-EfficientNet-B0. By adding the CBAM attention mechanism to the EfficientNet-B0 backbone network, the model's ability to focus on key regions is enhanced, thereby improving the classification performance of plant seedling images.

2. Fundamentals

2.1 Convolutional Neural Network

Convolutional neural networks (CNNs) are a typical model in deep learning and are widely used in various tasks such as image classification, object detection, semantic segmentation, and medical image analysis. Compared with traditional machine learning methods that rely heavily on manual design and handcrafted feature extraction, CNNs can automatically mine and learn deep features—such as edges, textures, and high-level semantics—from images in an end-to-end manner, using stacked convolutional and pooling layers. This bypasses the limitations of manual feature engineering, leading to stronger feature representation and generalization capabilities. Consequently, CNNs effectively improve the accuracy and robustness of various image recognition tasks.

The basic structure of a CNN mainly includes an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer. In the convolutional layers, convolution kernels slide over the input image to extract both shallow and deep features, capturing local details such as edges, textures, and geometric shapes. The pooling layers reduce the size of feature maps through down-sampling, which decreases the number of network parameters and computational cost, while also suppressing interference from irrelevant backgrounds and enhancing the model's translation invariance and robustness. The fully connected layers are responsible for integrating and mapping the multi-dimensional features extracted by previous layers, converting the feature information into class probabilities to complete the image classification task. The basic structure of the network is shown in Figure 1.

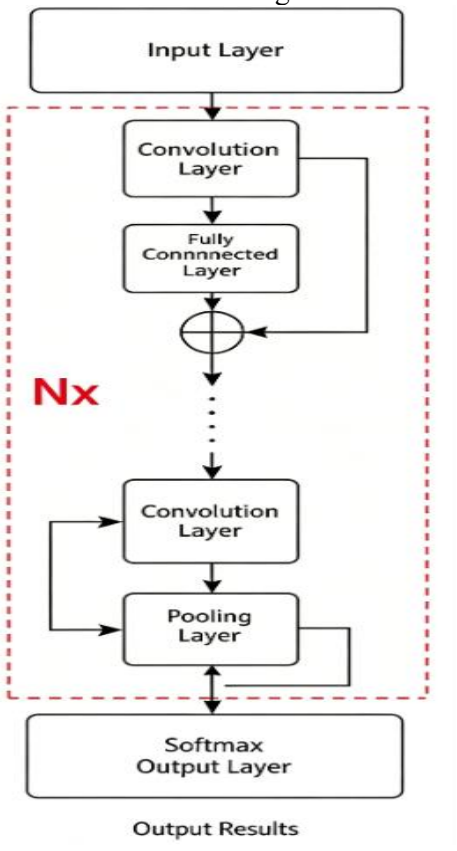

Figure 1. Architecture of the Network

2.2 EfficientNet Model

EfficientNet is a lightweight convolutional

neural network model proposed by Google [6]. Traditional convolutional networks usually improve performance by independently increasing network depth, width, or input resolution, but this approach tends to waste computational resources and increase model complexity. To address this issue, EfficientNet introduces a compound scaling method that uniformly coordinates the relationships among depth, width, and input resolution, thereby achieving a balance between model performance and computational efficiency.

EfficientNet's core MBConv block combines depthwise separable convolutions with residual connections, reducing parameters while enhancing feature extraction. Compared with traditional CNNs, EfficientNet achieves lower computational complexity and high accuracy, making it suitable for mobile and embedded devices [7]. In this paper, EfficientNet-B0 is adopted as the backbone a lightweight, efficient baseline for plant seedling recognition. Its compound scaling method enables higher performance with fewer parameters.

2.3 CBAM Attention Mechanism

CBAM is a lightweight attention mechanism module [8]. This module mainly consists of two components: channel attention and spatial attention. It helps the model focus more on critical feature regions, thereby improving the expressiveness of image features.

In the channel attention module, CBAM first applies global average pooling and max pooling to the input feature map respectively, then learns the importance weights of each channel through a multi-layer perceptron (MLP), and finally generates the channel attention map using a sigmoid function. Its computation process is as follows:

$$M_c(F) = \sigma\big(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))\big) \quad (1)$$

In the spatial attention module, CBAM compresses the channel dimension and then uses a convolution operation to generate the spatial attention map, allowing the model to focus more on key region information within the image. The calculation formula is as follows:

$$M_s(F) = \sigma\big(f^{7\times7}([AvgPool(F); MaxPool(F)])\big) \quad (2)$$

The CBAM module has a simple structure and a small number of parameters, making it easy to embed into different convolutional neural networks. Therefore, it has been widely applied in fields such as image classification, object detection, and medical image analysis [9].

In this paper, the CBAM attention mechanism is introduced on the basis of the EfficientNet-B0 model, aiming to enhance the model's ability to focus on key regional features of plant leaves, thereby improving the performance of plant seedling image classification.

3. Experimental Analysis

3.1 Dataset and Experimental Setup

The experiments in this paper are conducted using the Plant Seedlings Dataset for training and testing. This dataset contains several different plant seedling species, which are highly similar in color, texture, and leaf morphology, especially at the early seedling stage where inter-class differences are small. Therefore, to improve the model's generalization ability under complex environments, data augmentation techniques are applied to the dataset, including random flipping, random rotation, random cropping, and normalization [10]. At the same time, the model must not only extract features but also effectively focus on key regions of the plants to improve classification accuracy.

The dataset is divided into a training set (80%) and a validation set (20%). Table 1 lists the parameter settings for model training.

Table 1. Model Parameter Settings

Parameter	Setting
GPU	NVIDIA RTX 4060
Batch Size	32
Epoch	20
Learning Rate	0.0001
Input Size	224×224

In addition, the distribution of sample counts for each class is shown in Figure 2.

As can be seen from the figure 2, some classes have relatively many samples, while a few classes have fewer samples. Thus, the dataset suffers from a certain degree of class imbalance. If the model cannot effectively learn the features of under-represented classes, classification bias may occur. Therefore, we adopt a combination of data augmentation and attention mechanisms to improve the model's ability to learn features of different plant categories.

3.2 Model Construction

In this paper, EfficientNet-B0 is used as the

backbone network, and the CBAM attention mechanism is inserted after each MBConv block to enhance the model's ability to focus on critical feature regions. EfficientNet-B0 itself has few parameters, high training efficiency, and good classification performance, while the CBAM module further improves the model's ability to represent important regional features [11]. Figure 3 shows the overall architecture of the model.

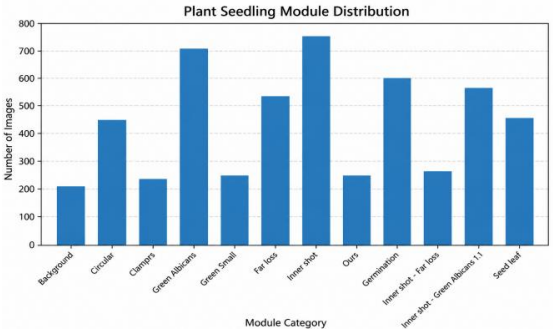

Figure 2. Sample Distribution

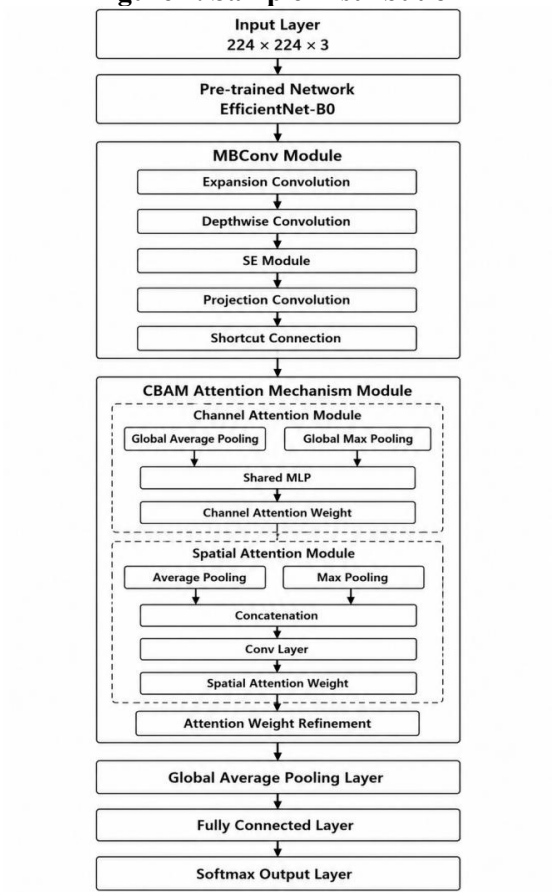

Figure 3. Model Architecture Diagram

The processing flow of the model is as follows: first, the convolutional layer extracts shallow texture features of plant images, and then the MBConv blocks further extract deep semantic features. To make the model pay more attention to key regions of plant leaves, the CBAM attention mechanism is added to the network. The CBAM module weights the feature maps from both channel and spatial dimensions, enabling the model to focus more on important regions such as leaf edges, texture, and color variations, thereby effectively reducing interference from background noise.

3.3 Analysis of Model Training Results

To intuitively observe the training process and performance of the model, this paper plots the loss and accuracy curves of the training and validation sets over epochs, as shown in Figure 4.

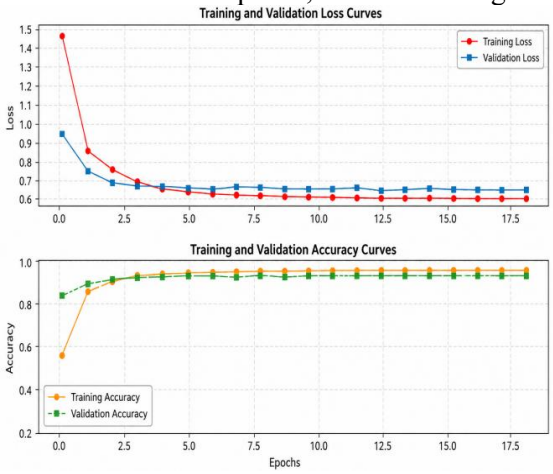

Figure 4. Loss and Accuracy Curves

From the loss curves, it can be seen that the loss value decreases rapidly at the beginning of training, indicating that the model quickly learns the basic features of plant seedling images. As the number of epochs increases, the loss values of both the training and validation sets gradually stabilize and finally converge smoothly, without significant fluctuations or upward trends. This demonstrates that the training process is stable, with no severe overfitting, and the model has good generalization ability.

From the accuracy curves, the model accuracy increases consistently with the number of epochs and gradually stabilizes in the later stage of training. The accuracy trends of the training and validation sets are generally consistent, and the final validation accuracy stabilizes at a relatively high level, indicating that the model effectively learns the key features of plant seedlings and achieves good classification performance. The accuracy increase slows down significantly in the later stage of training, suggesting that the model parameters are gradually approaching the optimal state, exhibiting good convergence performance and robustness.

3.4 Performance Evaluation and Analysis

To verify the effectiveness of the proposed model, this paper conducts comparative experiments between the CBAM-EfficientNet-B0 model and the baseline EfficientNet-B0, ResNet18, and MobileNetV2 models. The experimental results of different models are shown in Table 2.

Table 2. Comparison of Experimental Results of Different Models

modle	Accuracy
CBAM-EfficientNet-B0	0.969
EfficientNet-B0	0.789
ResNet18	0.724
MobileNetV2	0.712

From the table 2, it can be seen that the CBAM-EfficientNet-B0 model outperforms the other comparative models in terms of accuracy.

Compared with the baseline EfficientNet-B0 model, the improved model with the CBAM attention mechanism achieves significantly higher recognition accuracy. This fully demonstrates that the CBAM module effectively optimizes the feature selection and focusing capability of the original network, compensating for the baseline network's insufficient capture of key seedling features.

Compared with the ResNet18 model, the proposed model achieves a more substantial improvement in accuracy. This is because, although ResNet18 can extract deep semantic features, it is not sufficiently suitable for fine-grained classification tasks such as plant seedling recognition.

In addition, although MobileNetV2 has good lightweight characteristics, its network depth is relatively limited, which sometimes leads to deficiencies in extracting features from complex plant images, resulting in slightly lower classification performance than the proposed model.

In summary, the CBAM attention mechanism can effectively enhance the representational power of the EfficientNet-B0 model and improve the performance of seedling image recognition.

3.5 Visualization Results Analysis

To further verify the classification performance of the model, we plot several visualization results that make the model performance easier to interpret.

Figure 5 shows the confusion matrix results of the model on the test set.

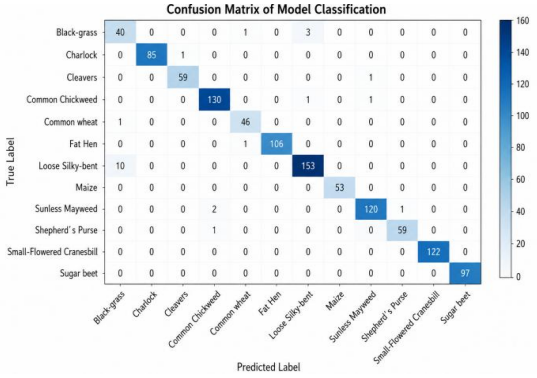

Figure 5. Confusion Matrix

As shown in Figure 5, the diagonal elements of the confusion matrix have higher values and darker colors than off-diagonal ones, indicating that most plant seedling samples are correctly classified. Only a few misclassifications occur among categories with very similar morphology, such as those with close leaf shapes and colors. This is because seedlings naturally share similar leaf shape, texture, and color at early growth stages—a common challenge in fine-grained classification. Overall, the proportion of misclassified samples is very low, and the model performs stably across all categories.

Figure 6 presents the recognition accuracy results for each category.

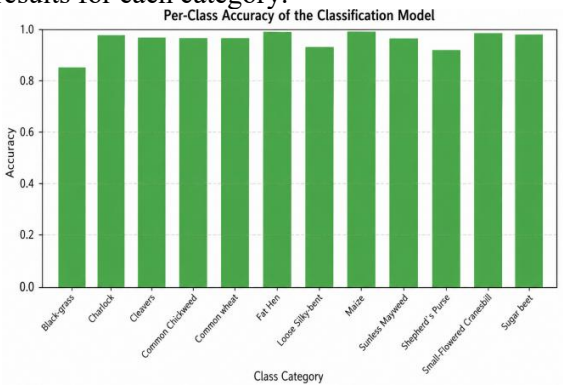

Figure 6. Recognition Accuracy per Category

Figure 7 shows the ROC curve results of the model.

It can be observed from Figure 6 that the recognition accuracy of the model for most plant seedling categories is close to 100%, with only a few categories being slightly lower, indicating excellent overall recognition performance. For the categories with lower accuracy, the main reason is that their leaf shapes and sizes are particularly similar to those of other categories, and the model is easily disturbed by background noise, illumination changes, and other factors during feature extraction, leading to insufficient feature discrimination. Nevertheless, from the overall distribution, the accuracy of all

categories remains at a high level, demonstrating that with the help of the CBAM attention mechanism, the proposed model can effectively capture the key distinguishing features among different seedling categories, exhibiting strong fine-grained discrimination ability.

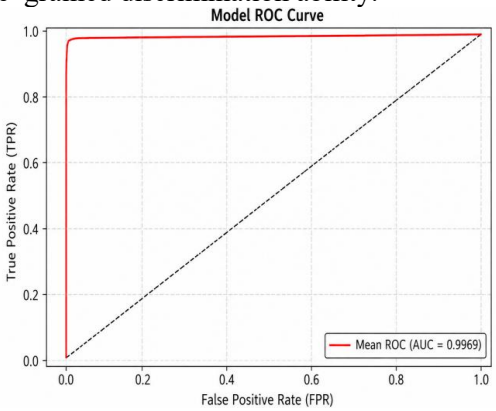

Figure 7. ROC Curve

The ROC curve intuitively reflects the recognition performance of the model under different classification thresholds. The closer the curve is to the upper-left corner and the closer the AUC value is to 1, the better the classification performance. As can be seen from Figure 7, the ROC curve of the proposed model is tightly aligned with the upper-left corner, with an AUC value close to 1, indicating that the model has a strong ability to distinguish between positive and negative samples and maintains stable recognition performance under different threshold conditions, demonstrating good robustness. This result further verifies the effectiveness of the CBAM attention mechanism in enhancing the model's feature extraction capability and improving classification stability.

In summary, the visualization results from multiple perspectives show that the CBAM-EfficientNet-B0 model achieves excellent recognition performance in the plant seedling image classification task. The per-category accuracy, ROC curve, PR curve, and model comparison results all demonstrate the effectiveness and superiority of the proposed method, while also indicating that the CBAM attention mechanism has significant application value in fine-grained plant recognition tasks.

4. Conclusion

To address the problems of small inter-class differences, complex background interference, and insufficient feature extraction capability of traditional convolutional neural networks in plant seedling image recognition, this paper proposes a method based on CBAM-EfficientNet-B0. This method uses EfficientNet-B0 as the backbone network and integrates the CBAM attention mechanism after each MBConv block. By combining channel and spatial attention, it effectively enhances the model's ability to focus on key regional features of plant leaves, thereby improving the classification performance of plant seedling images.

Experiments are conducted on the Plant Seedlings Dataset, and data augmentation is employed to improve the model's generalization ability. The experimental results show that the proposed CBAM-EfficientNet-B0 model outperforms ResNet18 and MobileNetV2 in terms of evaluation metrics. The model achieves a classification accuracy of over 96%, indicating that the addition of the CBAM attention mechanism enables the model to more effectively extract key features such as leaf texture, edges, and color.

Furthermore, the analysis of the loss and accuracy curves shows that the model exhibits good convergence performance and stability during training, with no obvious overfitting. In addition, visualization results including the confusion matrix and ROC curve further validate the effectiveness and robustness of the proposed model for plant seedling classification. The ROC curve lies close to the upper-left corner with an AUC value near 1, indicating strong classification capability.

In summary, the proposed model not only effectively improves the accuracy of plant seedling image recognition but also maintains good lightweight characteristics, offering practical application value. In smart agriculture scenarios, this method can be applied to automatic plant classification, intelligent agricultural monitoring, and vision-based agricultural robotics, contributing positively to the advancement of agricultural intelligence.

Although the proposed method achieves satisfactory experimental results, certain limitations remain. For example, the current experimental dataset is relatively limited in scale, and the model's adaptability to complex natural environments needs further improvement. Moreover, misclassifications may still occur for some plant categories with highly similar morphologies. Therefore, in future work, we will expand the scale of the plant image dataset and optimize the model using techniques such as

transfer learning, multi-scale feature fusion, and object detection. We will also consider deploying the model on mobile or embedded devices to enable real-time recognition and intelligent management of plant seedlings, thereby further enhancing the model's applicability in practical agricultural scenarios.

References

[1] DANG W Y, WANG Z L. Practical application of computer and IoT technology in smart agriculture. Southern Agricultural Machinery, 2026, 57(7): 170-173.

[2] ZHOU W Z, YANG T R, ZHONG Z H. Plant phenotypic feature image recognition system based on deep learning. Computer Knowledge and Technology, 2024, 20(36): 49-52. DOI: 10.14004/j.cnki.ckt.2024.1863.

[3] ZHOU Z H. Machine Learning. Beijing: Tsinghua University Press, 2016.

[4] QIU X P. Neural Networks and Deep Learning. Beijing: Publishing House of Electronics Industry, 2020.

[5] Tan M, Le V Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. CoRR, 2019, abs/1905.11946

[6] Shin M, Seo J, Lee B I, et al. Defective Photovoltaic Module Detection Using EfficientNet-B0 in the MachineVision Environment. Machines, 2026, 14 (2): 232-232. DOI: 10.3390/MACHINES14020232.

[7] Juseong L, Hoyoung T, Jongsun P. Energy Efficient Canny Edge Detector for Advanced Mobile Vision Applications. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28 (4): 1037-1046. DOI:10.1109/tcsvt.2016.2640038.

[8] Jiang P, Zhang J, Chen J. Enhanced rain removal network with convolutional block attention module (CBAM): a novel approach to image de-raining. EURASIP Journal on Advances in Signal Processing, 2025, 2025 (1): 9-9. DOI: 10.1186/S13634-025-01212-Z.

[9] YANG H K, ZHU B W, ZHANG Y M, et al. A review of plant disease image recognition algorithms based on deep learning. Application of Electronic Technique, 2025, 51(1): 1-7. DOI: 10.16157/j.issn.0258-7998.245285.

[10] WEI J A. Application of computer vision technology in image data processing. Popular Standardization, 2026, (5): 149-151.

[11] Yao J, Liu H, Ye J. Enhanced EfficientNet-B0 with Dual Attention Mechanisms for Food Category Classification in X-ray Images. Journal of Nondestructive Evaluation, 2026, 45 (2): 55-55. DOI: 10.1007/S10921-026-01348-4.