Pertanika J. Sci. & Technol. 34 (2): 1181 - 1202 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

FCA_Unet: An Enhanced UNet with Feature Coordination Attention for Wheat Leaf Disease Severity Assessment

Hongyan Zang[1,2], Annie Anak Joseph[1]*, Kuryati binti Kipli[1], David Bong Boon Liang[1], Wanzhen Wang[1,2], and Rong Liu[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
[2]College of Computer and Information Engineering, Qilu Institute of Technology, 250200 Jinan, China

ABSTRACT

Accurate monitoring of wheat leaf disease severity, critical for crop protection and precision agriculture, has attracted extensive attention in the context of smart agriculture. Existing methods face notable limitations: manual assessment is inefficient and subjective; classification-based approaches lack generalisability; segmentation models struggle with poor small-lesion detection, low boundary accuracy, and high parameter counts. To tackle these issues, we propose FCA_Unet, an enhanced UNet framework. The model consists of two key components: (1) the integration of the designed A_CBAM attention mechanism module and multi-scale feature fusion (MFFE) module. The attention module, based on deformable convolution, enhances focus on lesion-specific attributes while suppressing irrelevant background information. In contrast, the MFFE enables cross-layer fusion, integrating high-level semantic features for disease classification and low-level contextual features for boundary optimisation. (2) An optimised backbone network: The Inception multi-branch structure, coordinated attention mechanism (INCA), and MFFE are integrated into ResNet50 to reduce parameter redundancy while preserving feature extraction efficiency, thus meeting the lightweight deployment requirements of agricultural field equipment. Experimental results demonstrate that FCA_Unet achieves 89.85% mIoU on wheat stripe rust and powdery mildew, representing an 11.8% improvement over UNet, with a parameter count of 29.95M (68.2% of UNet's). Combined with disease severity indices, the model proves its superiority in both segmentation accuracy and severity assessment, providing robust support for automated wheat disease monitoring in practical agricultural scenarios.

Keywords: Attention mechanism, feature fusion, semantic segmentation, wheat leaf disease severity

ARTICLE INFO

Article history:
Received: 24 October 2025
Accepted: 19 December 2025
Published: 30 April 2026

DOI: https://doi.org/10.47836/pjst.34.2.24

E-mail addresses:
22010336@siswa.unimas.my (Hongyan Zang)
jannie@unimas.my (Annie Anak Joseph)
kkuryati@unimas.my (Kuryati binti Kipli)
bbldavid@unimas.my (David Bong Boon Liang)
22010329@siswa.unimas.my (Wanzhen Wang)
1274657903@qq.com (Rong Liu)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

INTRODUCTION

Wheat is a staple crop critical to global food security and agricultural development (Hassan et al., 2024; Kaur et al., 2023). These diseases spread rapidly through leaf lesions, if not detected in time, they can cause yield losses of 10% -70% (Kumar et al., 2024; Kumar & Kukreja, 2024). Accurate segmentation of leaf lesions is essential for early warning and targeted management (Sharma & Sethi, 2024; Shi et al., 2024; Zhang & Zhang, 2022).

Traditional approaches rely on manual inspection, which is time-consuming, subjective, and unsuitable for large-scale monitoring (Mi et al., 2020; Yi et al., 2024). With advances in computer vision and deep learning, image-based segmentation has become a promising tool for monitoring plant disease (Badrinarayanan et al., 2017; Hassan et al., 2024). UNet and its variants are widely used in this field due to their encoder-decoder structure and strong performance in pixel-level annotation (Rai & Pahuja, 2023; Shwetha et al., 2024; Wang et al., 2023; Yi et al., 2023; Zhang & Zhang, 2022; Zhang et al., 2022). Yet, their application in real field conditions remains challenging. Images often suffer from uneven lighting and cluttered backgrounds, while lesion appearance varies across disease stages (Liu et al., 2024; Wang et al., 2023). Moreover, agricultural monitoring frequently relies on drones or portable devices with limited computational resources, making conventional deep models too heavy for real-time deployment (Shi et al., 2024).

Recent advances in attention mechanisms, multi-scale feature fusion, and lightweight designs address these issues (Chen et al., 2021; Mishra et al., 2024; Yang et al., 2024; Yi et al., 2023), but balancing high accuracy with low complexity for practical field use remains an ongoing challenge. In this study, we improved the UNet model to balance segmentation accuracy and efficiency for wheat leaf disease detection. The model incorporates:

1. An A_CBAM attention mechanism for precise lesion location under complex backgrounds.
2. A multi-scale feature fusion module (MFFE) for integrating deep and shallow multi-scale features.
3. A lightweight FCA_ResNet backbone that reduces parameters while preserving feature extraction capacity.

Experiments on stripe rust and powdery mildew datasets demonstrate that FCA_Unet achieves higher accuracy than conventional models while maintaining efficiency suitable for edge deployment. This work provides an effective framework for automated and precise field monitoring of wheat diseases.

The remainder of this paper is organised as follows. Section 2 reviews related work. Section 3 presents the data material and the FCA_Unet model. Section 4 describes the results and analysis. Section 5 reviews the discussion and conclusion.

RELATED WORKS

In the development of smart agriculture, accurate monitoring of leaf diseases, particularly the precise estimation of disease severity, is critical for safeguarding crop yield and quality. Early research mainly adopted image classification approaches. For example, Hayit et al. (2021) proposed the Yellow-Rust-Xception model, which classifies wheat stripe rust into six severity levels with an accuracy of 91%. A model introduced by Nigam et al. (2022) for stem rust severity estimation achieved training and testing accuracies of 98.41% and 96.42%, respectively, with lower computational costs. Similarly, a C-DenseNet model incorporating convolutional block attention modules (CBAM) attained 97.99% accuracy on a six-level stripe rust grading dataset (Mi et al., 2020). While such classification methods achieve high accuracy, they provide only coarse severity estimates rather than precise quantification.

Segmentation-based methods have therefore attracted increasing attention. Semantic segmentation (Li et al., 2022; Zhao et al., 2025), instance segmentation (Sharma et al., 2024), and convolutional neural networks (CNNs) have shown strong potential in disease diagnosis and precision agriculture. In these approaches, lesion areas are segmented to estimate pixel counts or lesion area, which are then mapped to severity levels. For instance, Gao et al. (2023) combined simple linear iterative clustering super pixel segmentation with random forest classification, achieving an accuracy of 93.22% in stripe rust severity estimation. UNet, as a representative encoder-decoder model, has become a dominant choice in leaf disease segmentation due to its efficiency in pixel-level labelling. In stripe rust studies, UNet achieved an IoU of 0.9563 (Hassan et al., 2024). Sharma et al. (2024) integrated PointRend segmentation with EfficientNet classification, reporting an accuracy of 99.43%. Kumar et al. (2024) proposed a hybrid approach (FERSPNET-50) combining panoptic segmentation and object detection, outperforming YOLOv4 and RetinaNet in precision for rust detection.

Attention mechanisms and feature fusion have further advanced segmentation performance. These methods enable models to capture small lesions and irregular lesion boundaries more effectively. For example, Wang et al. (2023) proposed MFBP-UNet with multi-scale feature extraction and dynamic sparse attention, achieving a mean IoU of 86.15% in pear leaf disease segmentation. Rai and Pahuja (2023) improved UNet with attention gates, enhancing northern leaf blight segmentation in maize with 98.97% accuracy. Zhang et al. (2022) and Ge et al. (2023) employed an adaptive fusion module to complementarily integrate global and local attention features. This effectively reduces missed and false detections in multi-target scenarios. Zhang et al. (2023) apply the attention mechanism separately to two dimensions: semantic feature extraction and boundary detail optimisation. Via an attention-driven feature interaction module, the network achieves mutual enhancement between semantic information and boundary features.

Similarly, Yao et al. (2024) and Yi et al. (2024) introduced coordinate attention and dilated convolutions to boost lesion segmentation accuracy in apple and grape leaf diseases. These works demonstrate that combining multi-scale fusion and attention mechanisms enhances robustness in complex disease scenarios.

Lightweight model design has also emerged as a major trend to enable deployment on mobile and edge devices. For instance, WE-DeepLabV3+ employed MobileNetV2 with efficient channel attention, achieving an mIoU of 82.0% with only 5.1M parameters (Wang et al., 2024). Wang et al. (2021) introduced a lightweight UNet with MobileNetV3 and coordinate attention, reducing complexity while achieving 88.87% mIoU. Liu et al. (2024) developed StripeRustNet, a two-stage lightweight model for stripe rust, achieving 98.65% mIoU for leaf segmentation and 86.08% for lesion segmentation, while significantly reducing annotation time. Compared with large models, lightweight designs provide a better balance between accuracy and efficiency, making them more suitable for real-time agricultural monitoring in the field.

The analysis of existing research shows that in the study of various computer vision models, high accuracy and lightweight design are still the two most important indicators pursued by deep learning in practical applications. The balance between the two has always been essential to research content.

MATERIALS AND METHODS

Data

This study takes wheat disease as an example to illustrate the efficiency of our proposed method. The dataset consists of 652 images, including two parts. One part comes from a public wheat stripe rust dataset collected by the Ezhou Research Station of Huazhong Agricultural University in Hubei, China (Liu et al., 2024). The dataset consists of 5,013 high-resolution images with a resolution of 3,472 × 4,624 pixels (all images on a uniform background). There are 542 segmented images in the dataset. Each image is labelled as background, leaf, and diseased spot. Another part comes from a custom dataset captured under field conditions. 110 images of wheat powdery mildew are selected. These images are annotated using LabelMe, also according to the categories of background, leaf, and diseased spot. Image examples are shown in Figure 1. All 652 samples are randomly divided into a training set, validation set, and test set according to the ratio of 8:1:1. The specific division is shown in Table 1. The training set is used to learn and optimise the parameters of the model, the verification set is used to adjust the hyperparameters during the training process and monitor whether the model is overfitting, and the test set is used to evaluate the performance of the model in the final, ensuring the objectivity and reliability of the evaluation results.

Figure 1. Six samples of wheat leaf disease datasets: stripe rust (Liu et al., 2024) and powdery mildew (Jiang et al., 2022)

Table 1
Statistics of the dataset

Dataset	Training set	Validation set	Testing set	Total
Stripe rust	434	54	54	542
Powdery mildew	88	11	11	110
Total	522	349	179	652

Proposed FCA_Unet Model

This section introduces FCA_Unet, an enhanced variant of the traditional UNet designed to improve segmentation performance on wheat leaf diseases. FCA_Unet adopts an encoder-decoder architecture based on UNet. As illustrated in Figure 2, the encoder is composed of the improved FCA_ResNet, derived from ResNet50, together with the attention mechanism A_CBAM, while the decoder consists of up-sampling, deconvolution, and the feature fusion module MFFE. After image resizing, inputs are fed into the backbone FCA_ResNet. Feature maps from four different stages are concatenated with the up-sampled features through A_CBAM. By means of up-sampling, the decoder restores the feature maps extracted by the encoder to the resolution of the input image. Finally, decision-level fusion is achieved through the MFFE module, producing accurate segmentation results.

The overall structure of the INCA module is shown in Figure 3. The coordinate attention (CA) mechanism is a lightweight channel attention approach that overcomes the spatial feature limitations of SE attention and the long-range dependency shortcomings of CBAM. Given the widespread distribution of wheat disease images, a multi-scale feature extraction method is required to capture long-range dependencies. FCA_ResNet (Zang et al., 2025) integrates CA into multi-branch convolutional blocks to construct the INCA module, which effectively emphasises multi-scale information and lesion features.

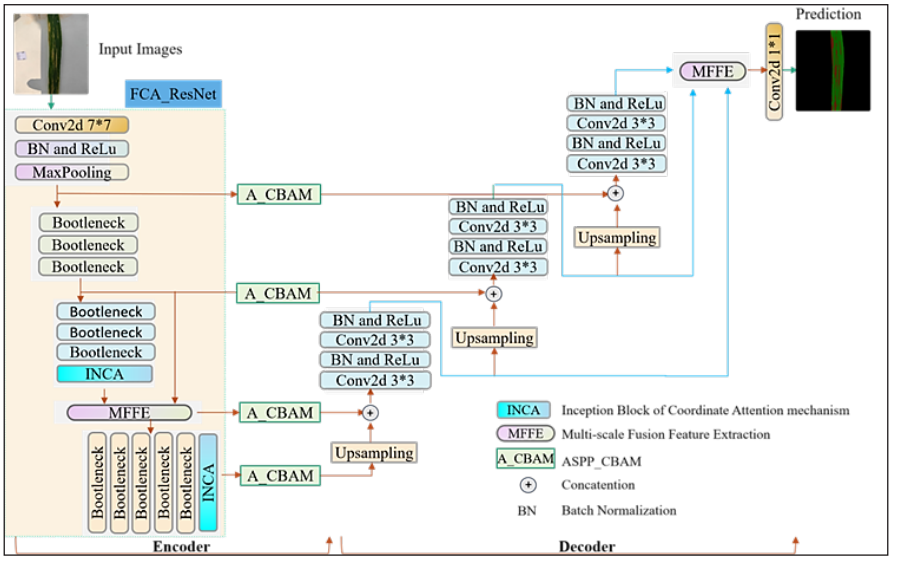


Figure 2. Structure overview of the proposed FCA_Unet segmentation model

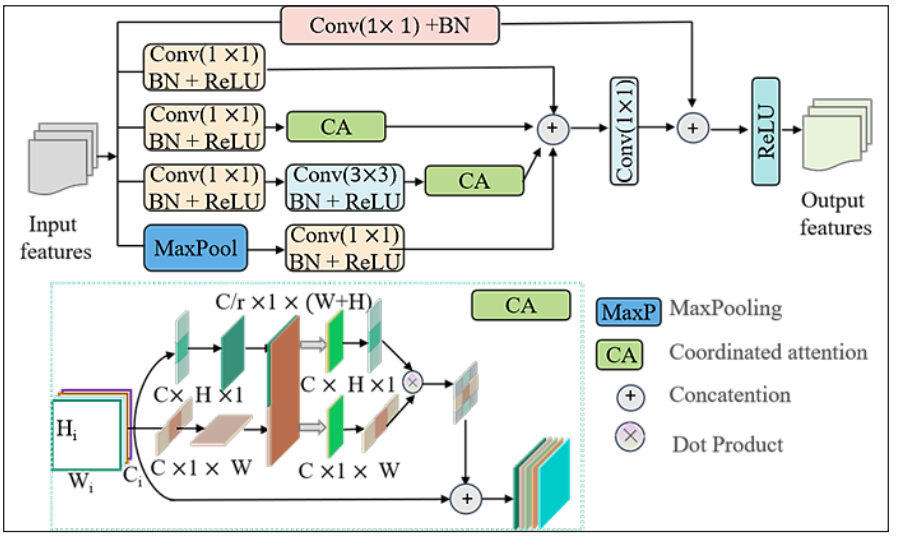


Figure 3. Structure diagram of the INCA module

Attention Mechanism Deform ASPP CBM (A_CBAM)

In standard convolution, kernels have fixed sizes and shapes, which may work well for objects with regular structures. However, in complex scenarios such as the segmentation of irregular disease spots on wheat leaves, deformable convolution offers greater potential than standard convolution. In the A_CBAM module, the 3×3 standard convolution within the ASPP is replaced by deformable convolution.

In a conventional convolution operation, each point on the output feature map corresponds to a sampling region of size K×K on the input feature map. In deformable convolution, each of these K×K sampling points learns an offset. As offsets are represented by coordinates, one output requires learning 2×K×K parameters. For an output of size H×W, this results in 2×K×K×H×W parameters in total, represented as the offset field in Figure 4, where N = K×K.

By allowing kernel positions to adjust dynamically, deformable convolution adapts the convolution weights to the spatial locations of sampling points. This enables the model to better accommodate non-uniform feature distributions in images, thereby enhancing its ability to capture complex structures, an advantage particularly critical in tasks such as object detection and semantic segmentation.

Multi-scale Feature Fusion Module (MFFE)

In segmentation tasks, the objective is to predict the category of each pixel. This requires not only robust representation of global target features but also precise preservation of boundary details, placing high demands on fine-grained feature extraction. Although data from different sources may contain redundancy, they also provide complementary information. In Figure 5, the MFFE module performs decision-level outputs from three different decoding stages, which enhances both the accuracy and the detail integrity of lesion segmentation. It further optimises the segmentation of small lesions and boundary regions.

The following section outlines the experimental results. It begins with providing information about the unique network configurations and related evaluation metrics. It then delves into conducting ablation experiments and comparative studies to analyse the performance of the proposed FCA_Unet segmentation model.

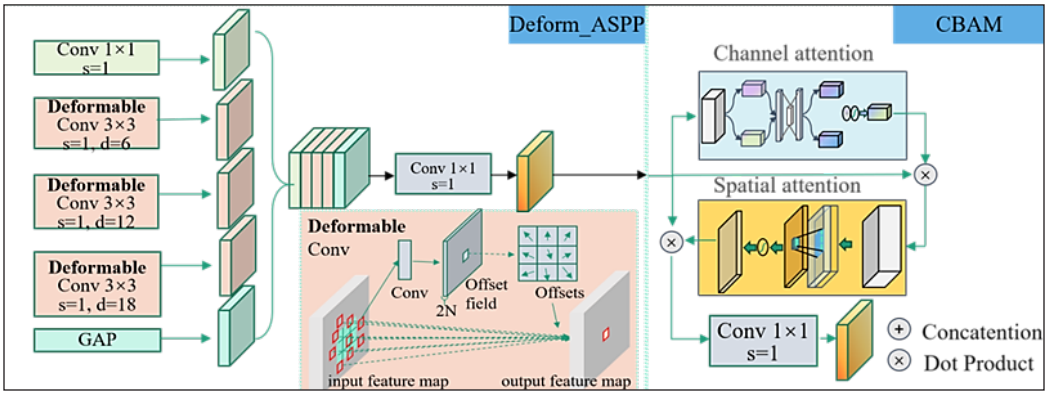


Figure 4. Structure diagram of A_CBAM module

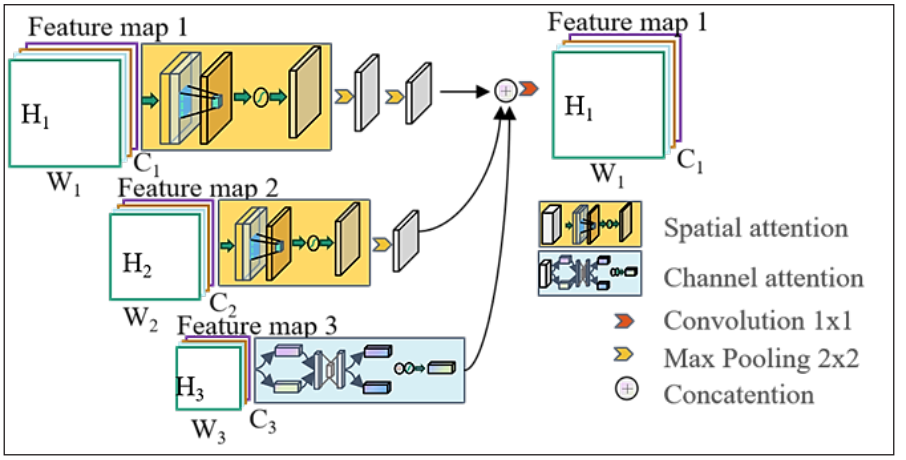


Figure 5. Structure diagram of the MFFE module

Evaluation Metrics

In the experiment, to quantitatively evaluate the performance of the proposed method and other comparison methods, mean pixel accuracy (mPA), mean intersection over union (mIoU), Disease Severity (DS), optimal model comprehensive score (OMCS), and other evaluation metrics were employed to assess the segmentation performance of the proposed model on wheat stripe rust.

Assuming there is a total of k classes (including a background class), P_{ij} denotes the number of pixels belonging to class i that are predicted as class j. Thus, P_{ii} represents the true positives (TP) and true negatives (TN), while P_{ij} and P_{ji} correspond to the false positives (FP) and false negatives (FN), respectively.

mPA

Pixel accuracy (PA) represents the ratio of correctly predicted pixels to the total number of pixels. Its calculation formula is given in Equation 1: -

$$\mathrm{PA}=\frac{\sum_{i=1}^{k} P_{ii}}{\sum_{i=1}^{k}\sum_{j=1}^{k} P_{ij}}=\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad [1]$$

mPA is defined as the mean value of the ratio between the number of correctly predicted pixels for each class and the total number of pixels of that class. The calculation formula is as shown in Equation 2: -

$$\mathrm{mPA}=\frac{1}{k+1}\sum_{i=1}^{k}\frac{P_{ii}}{\sum_{j=1}^{k} P_{ij}} \quad [2]$$

mIoU

The intersection over union (IoU) is the ratio of the intersection between the real and predicted sets to the union of the real and predicted sets for each category. The mIoU is a widely used evaluation metric in experimental studies of semantic segmentation and is calculated based on the average of the ratio of intersection and union of all categories. The calculation formulas are as shown in Equations 3 and 4: -

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP+FP+FN}} \quad [3]$$

$$\mathrm{mIoU}=\frac{1}{k}\sum_{i=1}^{k}\frac{\mathrm{TP}(i)}{\mathrm{TP}(i)+\mathrm{FP}(i)+\mathrm{FN}(i)} \quad [4]$$

DS

The DS index is used to quantify disease severity. DS value is the ratio of leaf spot area to leaf area, which is used to measure the severity of wheat leaf disease, as shown in Equation 5. This helps to assess the health status of the leaves in time and take appropriate measures.

$$\mathrm{DS}=\frac{S_{\mathrm{Disease}}}{S_{\mathrm{leaf}}}\times 100\% \quad [5]$$

OMCS

This paper proposes an OMCS. The core objective of OMCS is to identify the optimal model among multiple candidates that achieves both a high mIoU and a predicted DS value close to the ground-truth DS value.

For a single sample, the composite score of model *m* is defined as OMCS_m, as shown in Equation 6: -

$$\mathrm{OMCS}_m=\mathrm{NormmIoU}_m+(1-\mathrm{NormDS}_m) \quad [6]$$

Where: $\mathrm{NormmIoU}_m$ (as shown in Equation 7) denotes the normalised mIoU of model *m* (range: 0-1), and NormDS_m (as shown in Equation 8) represents the normalised DS of model *m* (range: 0-1).

$$\mathrm{NormmIoU}_m = \frac{\mathrm{mIoU}_m - \min(\mathrm{mIoU})}{\max(\mathrm{mIoU}) - \min(\mathrm{mIoU}) + \varepsilon} \quad [7]$$

mIoU_m: mIoU of model *m* on the current sample.

$\max(\mathrm{mIoU}), \min(\mathrm{mIoU})$: maximum and minimum mIoU among all models on the current sample.

ε: a very small value (e.g., 10^{-8}) to prevent division by zero.

$$\mathrm{NormDS}_m = \frac{\mathrm{DS}_m - \min(\mathrm{DS})}{\max(\mathrm{DS}) - \min(\mathrm{DS}) + \varepsilon} \quad [8]$$

$\max(\mathrm{DS}), \min(\mathrm{DS})$: the highest and lowest DS among all models for the current sample.

DS_m: as shown in Equation 9, the relative error between the predicted severity DS of model *m* and the ground truth is suitable for scenarios where the true DS values of different samples vary significantly. In scenarios where the true DS values are relatively small, absolute error can be used instead, i.e., $|\mathrm{PredDS}_m - \mathrm{TrueDS}|$.

$$\mathrm{DS}_m = \frac{|\mathrm{PredDS}_m - \mathrm{TrueDS}|}{\mathrm{TrueDS} + \varepsilon} \quad [9]$$

PredDS_m, TrueDS: the predicted DS of model *m* and the ground-truth DS label of the sample.

ε:a very small value (e.g., 10^{-8}) to avoid division by zero.

For each test sample, compare the comprehensive scores of all models. The model with the highest score is considered the optimal model for that sample, and its score is the OMCS of the sample. So, the best model can be expressed as Equation 10.

$$\text{Best Model} = \arg\max(\mathrm{OMCS}_m) \quad [10]$$

In addition, the number of parameters and FLOPs are also important indicators to evaluate the lightness of the model. They are used to determine whether the algorithm can be deployed on mobile devices.

RESULTS AND ANALYSIS

Ablation Experiment

To assess the effectiveness of the attention mechanism A_CBAM and the feature fusion module MFFE, we conducted comparative experiments and ablation studies. These two modules were embedded into different structures of the base model, and the relevant experimental comparison results are presented in Table 2.

By embedding the A_CBAM attention mechanism and MFFE feature fusion module into the base model and setting up different module combination schemes (such as embedding A_CBAM alone, embedding MFFE alone, and embedding both simultaneously), the impact of each module on model performance can be intuitively compared. The experimental results in Table 2 reflect the practical effects of A_CBAM in optimising feature attention and MFFE in multi-scale feature fusion from the perspective of quantitative metrics (e.g., mIoU, mAP, model parameter count, and inference time), providing data support for verifying the effectiveness of the modules.

Table 2
Evaluation of the effectiveness of the attention mechanism and feature fusion module

Number	Method	CBAM	A_CBAM	MFFE	mIoU	mAP	Parameter (MB)	Inference Time (ms)
1	UNet	×	×	×	0.7805	0.9744	43.93	**8.99**
2	FCA_Unet_1	×	×	×	0.7956	0.9751	39.57	11.63
3	FCA_Unet_2	√	×	×	0.7923	0.9780	39.64	13.57
4	FCA_Unet_3	×	√	×	0.8067	0.9781	**29.91**	20.08
5	FCA_Unet_4	×	×	√	0.7952	0.9782	39.61	14.41
6	FCA_Unet_5	×	√	√	**0.8077**	**0.9794**	29.95	20.68
7	FCA_Unet_6	√	×	√	0.8036	0.9738	39.68	14.57

Comparison with Baseline Model

When compared to the baseline model UNet, FCA_Unet_1, devoid of additional modules, outperformed UNet in both mIoU (0.7956 vs. 0.7805) and mAP (0.9751 vs. 0.9744), while reducing parameters by 9.9% (39.57 MB vs. 43.93 MB). This highlights the inherent superiority of FCA_Unet's basic architecture over the original UNet.

Impact of Individual Modules on Performance

FCA_Unet_3 with CBAM exhibited a marginal decrease in mIoU (0.7923 vs. 0.7956) but a modest improvement in mAP (0.9780 vs. 0.9751) compared to FCA_Unet_1 without CBAM.

However, integrating CBAM slightly increased parameters (39.64 MB vs. 39.57 MB) and significantly prolonged inference time (13.57 ms vs. 11.63 ms). This suggests CBAM did not effectively enhance the core segmentation metric (mIoU) under current experimental conditions, instead introducing additional computational overhead.

FCA_Unet_4 with A_CBAM achieved the highest mIoU (0.8067) among single-module models, representing a 1.11% improvement over FCA_Unet_1, with a slight mAP increase (0.9781 vs. 0.9751). Notably, A_CBAM substantially reduced model parameters (by ~24.4%, 29.91 MB vs. 39.57 MB), demonstrating its lightweight design advantage. Inference time increased significantly (20.08 ms vs. 11.63 ms), likely due to the computational costs of the self-attention mechanism.

FCA_Unet_5 with MFFE showed comparable mIoU to FCA_Unet_1 (0.7952 vs. 0.7956) but higher mAP (0.9782 vs. 0.9751). Parameters remained nearly unchanged (39.61 MB vs. 39.57 MB), though inference time increased (14.41 ms vs. 11.63 ms), indicating MFFE can improve classification accuracy (mAP) without increasing model size at the cost of slightly higher computational load.

Impact of Module Combinations

FCA_Unet_6, combining A_CBAM and MFFE, achieved the highest mIoU (0.8077) and mAP (0.9794) across all models, representing modest improvements over standalone A_CBAM (mIoU +0.10%, mAP +0.13%) and standalone MFFE (mIoU +1.25%, mAP +0.12%). With 29.95 MB parameters (like Model 4) and only a marginal inference time increase, the computational overhead is justified by performance gains. These results suggest A_CBAM and MFFE are complementary: A_CBAM enhances feature attention, while MFFE strengthens multi-scale feature integration.

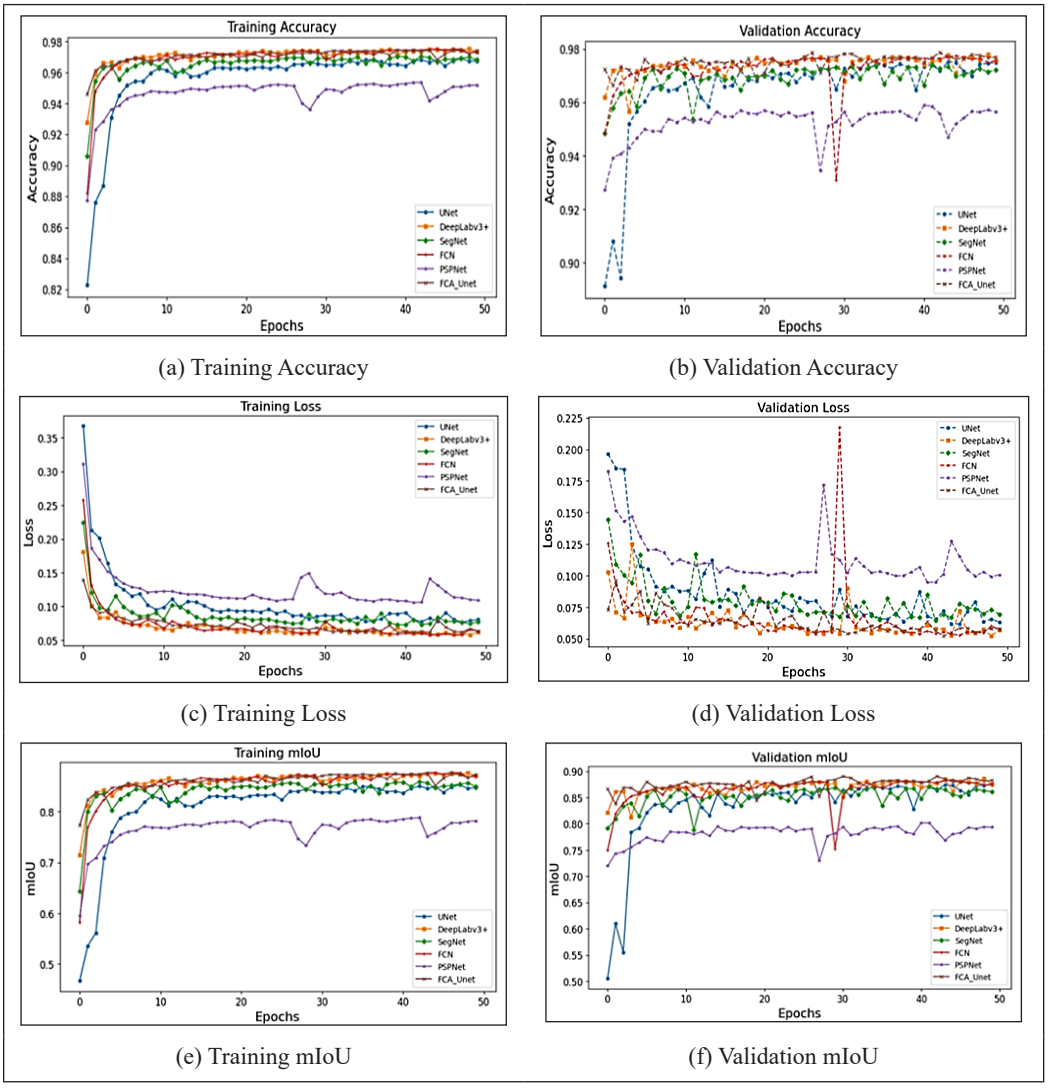
FCA_Unet_7, combining CBAM and MFE, achieved an mIoU of 0.8036, higher than standalone CBAM (0.7923) or MFFE (0.7952) but lower mIoU and mAP than the A_CBAM + MFFE combination. With parameters and inference times comparable to Models 3 and 5, this combination proved less effective in balancing performance and efficiency.

Experimental Comparison of Different Models

To demonstrate FCA_Unet's capability to capture feature information in wheat leaf disease images, four pre-trained segmentation models were compared. Comparative experiments used the same dataset under identical environmental conditions to ensure network performance comparability. Image size was fixed at 256×256 pixels for uniform processing, with a batch size of 4 and 50 epochs to avoid GPU memory issues and unstable training.

Training and Validation

This setup balanced efficiency and accuracy based on experimental results. A classic default learning rate of 0.001 was used to prevent slow convergence from low rates, suitable for different training stages. The Adam optimiser adjusted network parameters to minimise loss and accelerate convergence, facilitating deep learning model training. In (a) and (b) of Figure 6, the training accuracy and validation accuracy of FCA_Unet both remain around 97%. For the training loss and validation loss shown in (c) and (d), the losses of FCN and FCA_Unet are close, both reaching approximately 6%. In (e) and (f), FCA_Unet and FCN achieve the highest training mIoU, which can reach around 85%; however, in terms of validation mIoU, FCA_Unet and DeepLabv3+ perform the best, approaching 87%.



(a) Training Accuracy (b) Validation Accuracy

(c) Training Loss (d) Validation Loss

(e) Training mIoU (f) Validation mIoU

Figure 6. Comparative performance of different models in terms of accuracy, loss, and mIoU on the stripe rust training and validation sets

Testing

Differences in model performance can be visually assessed by randomly selecting several images for comparison. Comparative experiments between the proposed model and other approaches further confirm the efficacy and feasibility of the enhanced architecture. On the test set, the metrics recorded include mPA, mDice, mIoU, FLOPs, and model parameters. Here, mPA refers to the average pixel accuracy across all test images in the dataset.

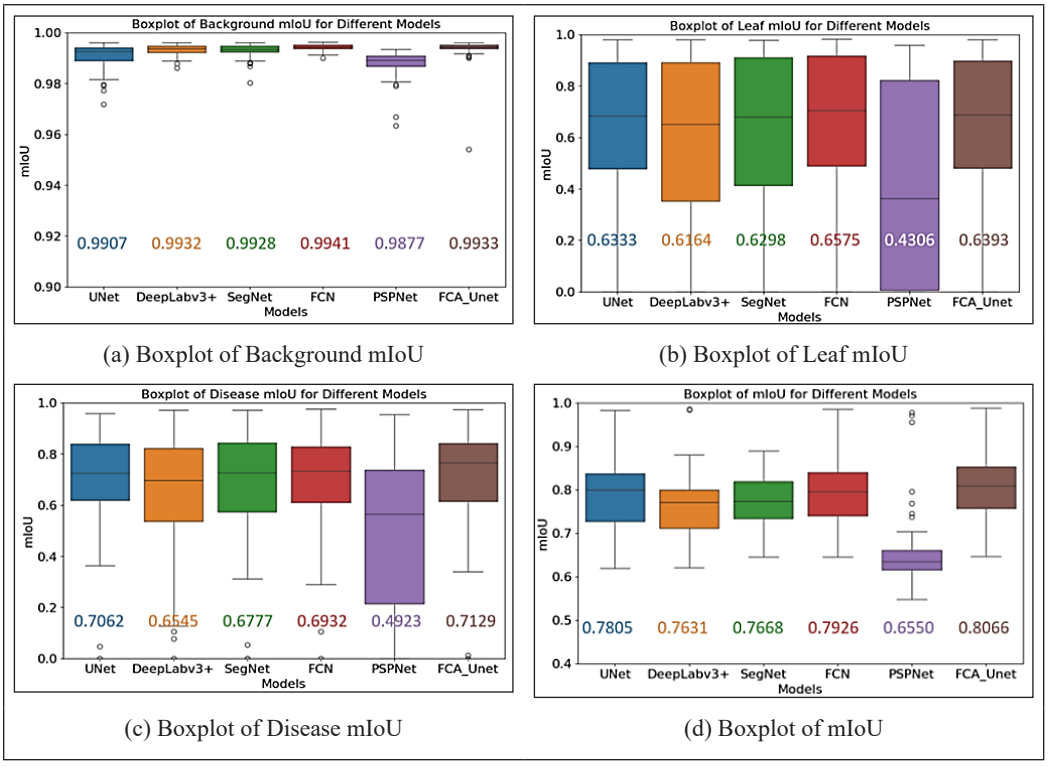
As Table 3 shows, FCA_Unet outperforms other models in mIoU, Dice, FLOPs, parameters, and best pixel accuracy. Notably, its top-1, top-5, and top-10 mIoU values across all tests are 89.85%, 88.54%, and 85.84%, respectively, demonstrating superior segmentation capability. The only exception is mAP, where FCA_Unet scores 97.57% second only to SegNet.

FCA_Unet's FLOPs reach 35.85G (55.69% higher than baseline UNet), but its parameter count is only 29.95M-68.18% of UNet's 43.93M, and comparable to FCN (23.52M), SegNet (29.44M), and PSPNet (15.18M). This efficiency stems from replacing UNet's ResNet50 backbone with FCA_ResNet, which has 10.76M parameters—57.9% fewer than ResNet50 (25.56M) (Zang et al., 2025).

Table 3
Performance comparison between FCA_Unet and mainstream segmentation models on the stripe rust test set

Model	Backbone	mPA	mIoU			FLOPs	Parameters	Best Pixel Accuracy
			Top 1	Top 5	Top 10			
UNet	ResNet50	97.44%	88.98%	85.81%	84.13%	23.02G	43.93M	99.56%
DeepLabv3+	ResNet50	97.42%	87.96%	82.21%	81.03%	45.77G	40.35M	99.65%
FCN	ResNet50	**97.76%**	89.60%	86.72%	84.20%	**5.55G**	23.52M	99.63%
SegNet	VGG16	97.59%	88.92%	83.56%	82.86%	40.14G	29.44M	99.59%
PSPNet	ResNet50	95.50%	88.19%	85.26%	83.20%	15.05G	**15.18M**	99.64%
FCA_Unet	ResNet_F_INCA	97.57%	**89.85%**	**88.54%**	**85.84%**	35.84G	29.95M	99.66%

Figure 7 compares IoU performance for wheat leaf disease spots. In background segmentation, FCN achieves the highest mean IoU (0.9941), followed by FCA_Unet and DeepLabv3+, with PSPNet lowest (0.9878). For leaf segmentation, PSPNet fluctuates significantly while FCA_Unet stays stable. FCA_Unet also leads in lesion segmentation. Overall, PSPNet is the weakest and least stable; FCN excels in background segmentation; SegNet shows consistent performance; and FCA_Unet delivers strong overall results with potential for greater stability.



(a) Boxplot of Background mIoU

(b) Boxplot of Leaf mIoU

(c) Boxplot of Disease mIoU

(d) Boxplot of mIoU

Figure 7. The performance of different models on the test set regarding the IoU and mIoU of the background, leaves, and disease spots

Visualisation

To provide a more intuitive comparison of segmentation results across different networks, we randomly selected wheat leaf disease images. The segmentation results for stripe rust and powdery mildew on the test set are shown in Figures 8 and 9, respectively, illustrating the performance of different network architectures. The first column displays the original wheat leaf images, the second column shows the manually annotated masks, and each subsequent column represents the outputs of different models. The mIoU and DS values for each sample under different models are summarised in Tables 4 and 5. As shown in Table 4, the proposed FCA_Unet consistently achieved the highest mIoU across both disease types, particularly on stripe rust samples.

Compared with other models, FCA_Unet produced more complete lesion segmentation, finer boundaries, and higher-quality masks. These gains likely result from its dilated convolutions and attention mechanisms, which capture contextual and small-lesion features, and from a lightweight fusion module in the decoder that refines boundaries. Its superior mIoU further confirms robustness and effectiveness in wheat leaf disease severity assessment.

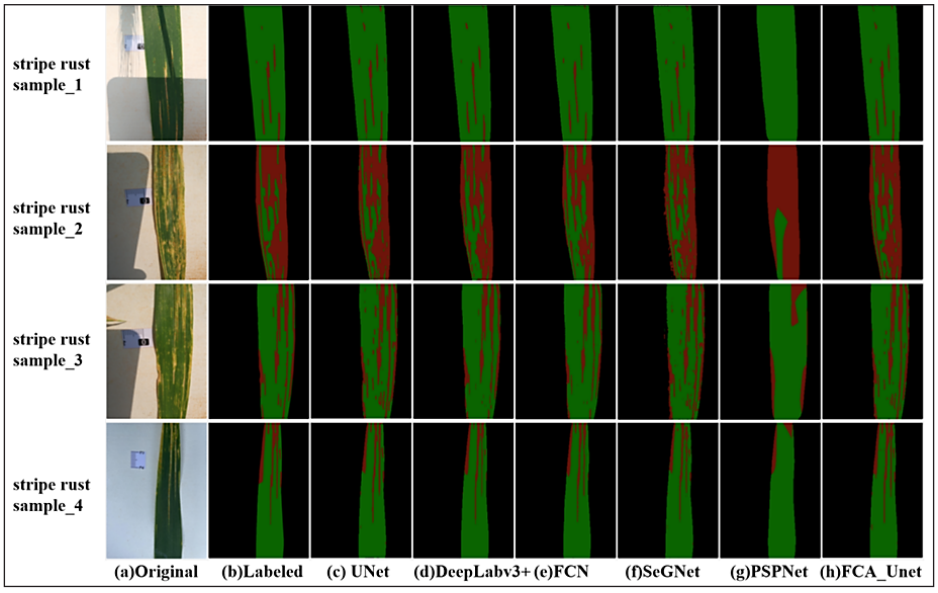


Figure 8. Example images of stripe rust demonstrating visual results obtained by segmentation models. Masks for background, normal leaf and disease spots are represented by black, green and red, respectively

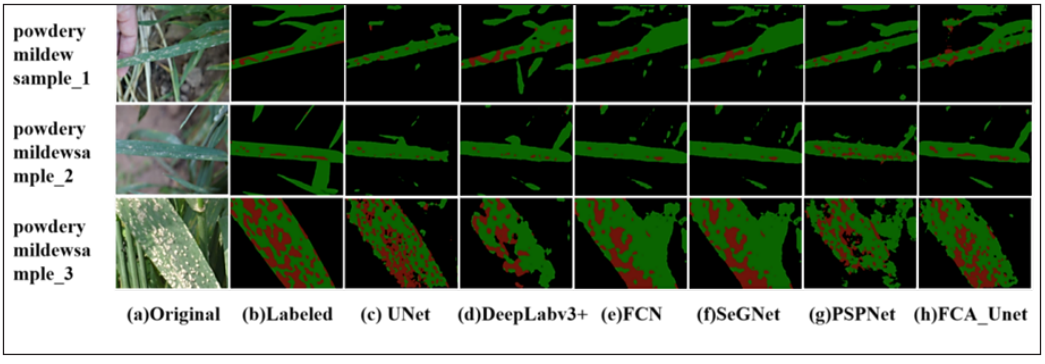


Figure 9. Example images of powdery mildew demonstrating visual results obtained by segmentation models. Masks for background, normal leaf and disease spots are represented by black, green and red, respectively

From Table 5, it can be said that different models produce markedly different DS predictions. For instance, in the case of stripe rust sample_1, the ground truth DS is 5.62%. The FCA_Unet predicts a DS of 5.86%, with an absolute error of only 0.24%. In contrast, the predictions from UNet, DeepLabv3+, FCN, SegNet, and PSPNet are 6.48%, 5.05%, 4.90%, 5.27%, and 0%, with absolute errors of 0.86%, -0.57%, -0.72%, -0.35%, and -5.62%, respectively.

Since this study focuses on disease severity, DS accuracy is crucial. However, Tables 4 and 5 show that a DS value close to the ground truth does not necessarily correspond to a high mIoU.

Table 4
Comparison of mIoU performance of different segmentation models on the test set

Number	Test sample	UNet	DeepLabv3+	FCN	SeGNet	PSPNet	FCA_Unet
1	stripe rust sample_1	0.8898	0.8796	0.896	0.8892	0.632	**0.8984**
2	stripe rust sample_2	0.8557	0.8103	0.8039	0.8351	0.6866	**0.8725**
3	stripe rust sample_3	0.8581	0.849	0.8833	0.8742	0.6603	**0.8852**
4	stripe rust sample_4	0.8824	0.8717	0.8774	0.8632	0.7041	**0.8985**
5	powdery mildew sample_1	0.4815	0.6176	0.6031	0.6012	0.4523	**0.624**
6	powdery mildew sample_2	0.5094	0.5737	0.6400	0.5512	0.4129	**0.6413**
7	powdery mildew sample_3	0.6796	0.5706	0.5999	0.4813	0.5936	0.7307

Table 5
Comparison of Dice Score (DS) performance of different segmentation models on the test set

Number	Test sample	TRUE	UNet	DeepLabv3+	FCN	SeGNet	PSPNet	FCA_Unet
1	stripe rust sample_1	5.62%	6.48%	5.05%	4.90%	5.27%	0.00%	**5.86%**
2	stripe rust sample_2	73.81%	74.17%	68.02%	61.66%	68.86%	86.54%	**72.10%**
3	stripe rust sample_3	21.99%	27.53%	**21.75%**	19.95%	26.96%	14.58%	24.84%
4	stripe rust sample_4	17.19%	15.37%	14.00%	14.13%	13.50%	7.13%	**16.84%**
5	powdery mildew sample_1	8.92%	5.80%	15.65%	**9.79%**	5.51%	2.14%	11.29%
6	powdery mildew sample_2	5.10%	2.82%	4.03%	8.79%	13.17%	0.00%	**6.13%**
7	powdery mildew sample_3	41.88%	55.27%	**38.29%**	32.50%	26.90%	28.69%	33.68%

A single DS value can be unreliable because DS is a relative measure: it only reflects the agreement between numerical proportions but cannot determine whether each diseased or healthy pixel is correctly predicted. By contrast, mIoU evaluates pixel-level agreement and thus provides a more rigorous measure.

To address this, we propose the OMCS, which integrates both mIoU and DS into a single, more intuitive indicator for identifying the best-performing model. The OMCS results derived from Tables 4 and 5 are illustrated in Figure 10. Here, the normalised ranges of Equations 8 and 9 are both in 0-1. Therefore, the OMCS in this case ranges from 0 to 2, where a larger value indicates better model performance. As shown in Figure 10, FCA_Unet is the optimal model in 5 out of 7 samples, demonstrating that FCA_Unet is the best-performing model among all those compared.

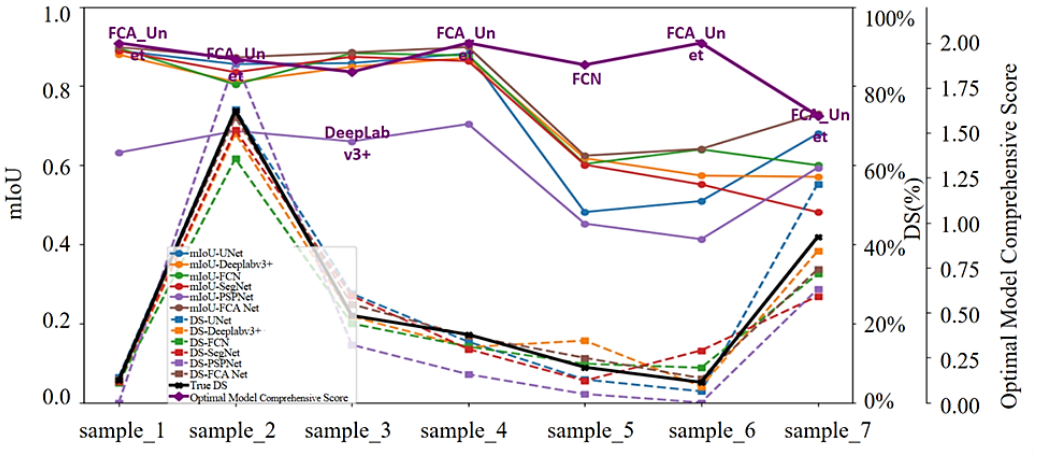


Figure 10. Comprehensive score analysis of the optimal model

Discussion

This study introduces FCA_Unet, a lightweight and high-performance model for wheat leaf disease segmentation. By integrating the A_CBAM attention mechanism and MFFE feature fusion module with an optimised FCA_ResNet backbone, the model achieves superior lesion segmentation accuracy and detail preservation while reducing parameters by 31.8% compared with UNet. Experimental results demonstrate its advantage over benchmark models, reaching a maximum mIoU of 89.85% and ranking as the optimal model in most test samples through OMCS evaluation.

Comparative experiments show that FCA_Unet achieves the highest mIoU of 89.85% on the test set, outperforming classical models such as FCN, PSPNet, and SegNet. Although FLOPs increased by 55.69% compared to UNet due to attention and dilated convolutions, the reduced parameter size facilitates deployment on resource-limited devices, a trade-off that aligns with the practical needs of agricultural on-site detection (Liu et al., 2025). To avoid the high computational cost incurred by traditional attention mechanisms, Al-Gaashani et al. (2025) introduced the multi-scale convolutional PoolFormer block to enhance the model's ability to capture both local and global dependencies.

In this paper, the OMCS, which integrates mIoU and DS, FCA_Unet demonstrates balanced advantages in both pixel-level segmentation and disease quantification, addressing the limitations of relying solely on DS. Notably, discrepancies observed between DS predictions and mIoU highlight that accurate severity assessment depends on precise pixel-level segmentation rather than ratio-based measures alone. By enhancing boundary delineation and small-lesion detection, FCA_Unet provides a more reliable foundation for disease severity evaluation, offering practical value for precision pesticide application in agriculture.

CONCLUSION

This study proposes the FCA_Unet model, an improved variant of UNet, which significantly enhances segmentation performance for wheat leaf diseases. The model's superiority arises from two main innovations: (1) the integration of the A_CBAM attention mechanism and MFFE fusion module, which improve lesion segmentation by capturing contextual information and refining multi-scale features, particularly for small targets and boundary details; and (2) the adoption of FCA_ResNet as the backbone, reducing parameters to 42.1% of ResNet50 and 68.2% of baseline UNet, thereby achieving model lightweighting without compromising accuracy. These findings highlight FCA_Unet as an efficient solution for automated wheat disease monitoring, with potential for broader application to other crop disease segmentation tasks.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Engineering, Universiti Malaysia Sarawak (UNIMAS), Qilu Institute of Technology, and the StripeRust-Pocket dataset for providing partial support for our experimental data.

REFERENCES

Al-Gaashani, M. S. A. M., Alkanhel, R., Ali, M. A. S., Muthanna, M. S. A., Aziz, A., & Muthanna, A. (2025). MSCPNET: A multi-scale convolutional pooling network for maize disease classification. *IEEE Access, 13*, 11423-11446. https://doi.org/10.1109/ACCESS.2024.3524729

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

Chen, S., Zhang, K., Zhao, Y., Sun, Y., Ban, W., Chen, Y., Zhuang, H., Zhang, X., Liu, J., & Yang, T. (2021). An approach for rice bacterial leaf streak disease segmentation and disease severity estimation. *Agriculture, 11*(5), Article 420. https://doi.org/10.3390/agriculture11050420

Gao, R., Jin, F., Ji, M., & Zuo, Y. (2023). Research on the method of identifying the severity of wheat stripe rust based on machine vision. *Agriculture, 13*(12), Article 2187. https://doi.org/10.3390/agriculture13122187

Ge, Y., Zhang, Q., Xiang, T., Zhang, C., & Bi, H. (2023). TCNet: Co-salient object detection via parallel interaction of transformers and CNNs. *IEEE Transactions on Circuits and Systems for Video Technology, 33*(6), 2600-2615. https://doi.org/10.1109/TCSVT.2022.3225865

Hassan, A., Mumtaz, R., Mahmood, Z., Fayyaz, M., & Naeem, M. K. (2024). Wheat leaf localisation and segmentation for yellow rust disease detection in complex natural backgrounds. *Alexandria Engineering Journal, 107*, 786-798. https://doi.org/10.1016/j.aej.2024.09.018

Hayit, T., Erbay, H., Varçın, F., Hayit, F., & Akci, N. (2021). Determination of the severity level of yellow rust disease in wheat by using convolutional neural networks. *Journal of Plant Pathology, 103*(3), 923-934. https://doi.org/10.1007/s42161-021-00886-2

Jiang, J., Liu, H., Zhao, C., He, C., Ma, J., Cheng, T., Zhu, Y., Cao, W., & Yao, X. (2022). Evaluation of diverse convolutional neural networks and training strategies for wheat leaf disease identification with field-acquired photographs. *Remote Sensing, 14*(14), Article 3446. https://doi.org/10.3390/rs14143446

Kaur, P., Harnal, S., Gautam, V., Singh, M. P., & Singh, S. P. (2023). Performance analysis of segmentation models to detect leaf diseases in tomato plant. *Multimedia Tools and Applications, 83*(6), 16019-16043. https://doi.org/10.1007/s11042-023-16238-4

Kumar, D., & Kukreja, V. (2024). Image segmentation, classification, and recognition methods for wheat diseases: Two decades' systematic literature review. *Computers and Electronics in Agriculture, 221*, Article 109005. https://doi.org/10.1016/j.compag.2024.109005

Kumar, D., Kukreja, V., & Singh, A. (2024). A novel hybrid segmentation technique for identification of wheat rust diseases. *Multimedia Tools and Applications, 83*(29), 72221-72251. https://doi.org/10.1007/s11042-024-18463-x

Li, Y., Qiao, T., Leng, W., Jiao, W., Luo, J., Lv, Y., Tong, Y., Mei, X., Li, H., Hu, Q., & Yao, Q. (2022). Semantic segmentation of wheat stripe rust images using deep learning. *Agronomy, 12*(12), Article 2933. https://doi.org/10.3390/agronomy12122933

Liu, K., Li, L., & Zhang, X. (2025). Fast and accurate identification of kiwifruit diseases using a lightweight convolutional neural network architecture. *IEEE Access, 13*, 84826-84843. https://doi.org/10.1109/ACCESS.2025.3564355

Liu, W., Chen, Y., Lu, Z., Lu, X., Wu, Z., Zheng, Z., Suo, Y., Lan, C., & Yuan, X. (2024). StripeRust-Pocket: A mobile-based deep learning application for efficient disease severity assessment of wheat stripe rust. *Plant Phenomics, 6*, Article 0201. https://doi.org/10.34133/plantphenomics.0201

Mi, Z., Zhang, X., Su, J., Han, D., & Su, B. (2020). Wheat stripe rust grading by deep learning with attention mechanism and images from mobile devices. *Frontiers in Plant Science, 11*, Article 558126. https://doi.org/10.3389/fpls.2020.558126

Mishra, M., Choudhury, P., & Pati, B. (2024). IoT-enabled plant leaf disease segmentation and multi-classification using mayfly bald eagle optimisation-enabled machine learning. *Multimedia Tools and Applications, 83*(21), 59747-59781. https://doi.org/10.1007/s11042-023-17680-0

Nigam, S., Jain, R., Prakash, S., Marwaha, S., Arora, A., Singh, V. K., Singh, A. K., & Prakasha, T. L. (2022). Wheat disease severity estimation: A deep learning approach. In R. Misra, N. Kesswani, M. Rajarajan, B.

Veeravalli, & A. Patel (Eds.), *Lecture Notes in Networks and Systems* (Vol. 340, pp. 185-193). Springer. https://doi.org/10.1007/978-3-030-94507-7_18

Rai, C. K., & Pahuja, R. (2023). Northern maize leaf blight disease detection and segmentation using deep convolution neural networks. *Multimedia Tools and Applications, 83*(7), 19415-19432. https://doi.org/10.1007/s11042-023-16398-3

Sharma, J., Kumar, D., Chattopadhay, S., Kukreja, V., & Verma, A. (2024). Automated detection of wheat powdery mildew using YOLACT instance segmentation. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimisation (Trends and Future Directions)* (pp. 1-4). https://doi.org/10.1109/ICRITO61523.2024.10522394

Sharma, T., & Sethi, G. K. (2024). Improving wheat leaf disease image classification with point rend segmentation technique. *SN Computer Science, 5*(2), Article 224. https://doi.org/10.1007/s42979-023-02571-w

Shi, L., Liu, Z., Yang, C., Lei, J., Wang, Q., Yin, F., & Wang, J. (2024). Lightweight U-Net-based method for estimating the severity of wheat fusarium head blight. *Agriculture, 14*(6), Article 938. https://doi.org/10.3390/agriculture14060938

Shwetha, V., Bhagwat, A., Laxmi, V., & Shrivastava, S. (2024). A custom backbone UNet framework with DCGAN augmentation for efficient segmentation of leaf spot diseases in jasmine plant. *Journal of Computer Networks and Communications, 2024*, Article 15. https://doi.org/10.1155/2024/5057538

Wang, C., Du, P., Wu, H., Li, J., Zhao, C., & Zhu, H. (2021). A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. *Computers and Electronics in Agriculture, 189*, Article 106373. https://doi.org/10.1016/j.compag.2021.106373

Wang, H., Ding, J., He, S., Feng, C., Zhang, C., Fan, G., Wu, Y., & Zhang, Y. (2023). MFBP-UNET: A network for pear leaf disease segmentation in natural agricultural environments. *Plants, 12*(18), Article 3209. https://doi.org/10.3390/plants12183209

Wang, Z., Yang, L., Wang, R., Lei, L., Ding, H., & Yang, Q. (2024). WE-DeepLabV3+: A lightweight segmentation model for *Panax notoginseng* leaf diseases. *Computers and Electronics in Agriculture, 227*, Article 109612. https://doi.org/10.1016/j.compag.2024.109612

Yang, T., Wang, Y., & Lian, J. (2024). Plant diseased lesion image segmentation and recognition based on improved multi-scale attention net. *Applied Sciences, 14*(5), Article 1716. https://doi.org/10.3390/app14051716

Yao, C., Yang, Z., Li, P., Liang, Y., Fan, Y., Luo, J., Jiang, C., & Mu, J. (2024). Two-stage detection algorithm for plum leaf disease and severity assessment based on deep learning. *Agronomy, 14*(7), Article 1589. https://doi.org/10.3390/agronomy14071589

Yi, X., Wang, J., Wu, P., Wang, G., Mo, L., Lou, X., Liang, H., Huang, H., Lin, E., Maponde, B. T., & Lv, C. (2023). AC-UNet: An improved UNet-based method for stem and leaf segmentation in *Betula luminifera*. *Frontiers in Plant Science, 14*, Article 1268098. https://doi.org/10.3389/fpls.2023.1268098

Yi, X., Zhou, Y., Wu, P., Wang, G., Mo, L., Chola, M., Fu, X., & Qian, P. (2024). U-Net with coordinate attention and VGGNet: A grape image segmentation algorithm based on fusion pyramid pooling and the dual-attention mechanism. *Agronomy, 14*(5), Article 925. https://doi.org/10.3390/agronomy14050925

Zang, H., Joseph, A. A., Zhang, S., Fu, H., Huang, L., & Huang, Z. (2025). FCA-ResNet: An improved model with enhanced multi-scale feature fusion and coordinate attention for wheat leaf disease classification. *International Journal of Engineering and Technology Innovation, 15*(2), 195-209. https://doi.org/10.46604/ijeti.2024.14304

Zhang, Q., Ge, Y., Zhang, C., & Bi, H. (2022). TPRNet: Camouflaged object detection via transformer-induced progressive refinement network. *The Visual Computer, 39*, 4593-460. https://doi.org/10.1007/s00371-022-02611-1

Zhang, Q., Sun, X., Chen, Y., Ge, Y., & Bi, H. (2023). Attention-induced semantic and boundary interaction network for camouflaged object detection. *Computer Vision and Image Understanding, 233*, Article 103719. https://doi.org/10.1016/j.cviu.2023.103719

Zhang, S., & Zhang, C. (2022). Modified U-Net for plant diseased leaf image segmentation. *Computers and Electronics in Agriculture, 204*, Article 107511. https://doi.org/10.1016/j.compag.2022.107511

Zhao, C., Li, C., Wang, X., Wu, X., Du, Y., Chai, H., Cai, T., Xiang, H., & Jiao, Y. (2025). Plant disease segmentation networks for fast automatic severity estimation under natural field scenarios. *Agriculture, 15*(6), Article 583. https://doi.org/10.3390/agriculture15060583