



Research Article

<https://doi.org/10.1631/jzus.B2500693>

Improving RGB image recognition in the YOLO11n algorithm for accurate detection of tea plant diseases

Jinxian TAO¹, Xiaoli LI¹✉, Jingfei ZHANG³✉, Muhammad SHOAIB¹, Muhammad adnan ISLAM¹, Ibrar AHMAD¹, Yong HE¹, Sitan YE¹, Yujie WANG¹, Binhui LIAO³, Mostafa GOUDA^{1,2}

¹College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China

²Department of Nutrition & Food Science, National Research Centre, Dokki, 12622, Giza, Egypt

³Liandu Agriculture and Rural Bureau, Lishui 323000, China

Abstract: Tea diseases, including brown and grey blight, result in significant yield and quality losses, especially in Longjing tea production. Traditional detection methods are prone to errors, while existing deep learning models often struggle to be robust in natural field conditions. To address these challenges, an improved lightweight detection model, Asymmetric Multi-level mechanism, Dynamic Snake convolution, Siou loss function-YOLO (ADS-YOLO), was developed and validated. In the method, a dataset comprising 5694 smartphone-captured images of tea leaves was established under natural lighting. Enhancements were implemented in the YOLO11n baseline algorithm through the incorporation of the Siou loss function for better bounding box regression, dynamic snake convolution (DSC), which realizes adaptive feature extraction based on the dynamic spatial context, and an Asymmetric Multi-level (AML) mechanism, which achieves lightweight feature fusion via adaptive multi-scale design. The results showed that ADS-YOLO achieved a precision of 0.935 and a recall of 0.870, compared to 0.894 and 0.818, respectively, using the baseline YOLO11n. Importantly, ADS-YOLO demonstrated a real-time performance of 137.1 FPS, coupled with reduced computational costs. ADS-YOLO improved mAP@0.5 by 6.4% compared with YOLOv5n and achieved up to 44.6% higher accuracy than YOLOv7t. In conclusion, ADS-YOLO achieved high accuracy, providing a scalable solution for real-time crop health monitoring and sustainable, precision agriculture for tea production.

Key words: Tea disease detection; YOLO11n; Convolutional module; Attention mechanism; Loss function.

1 Introduction

Globally, tea (*Camellia sinensis* L.) plays a vital role in both cultural heritage and the economy. Especially in China, tea is among the most valuable crops, with over 3.2 million hectares dedicated to its cultivation (Liu et al. 2024). However, tea plantations face serious threats from several diseases, notably tea algal leaf spot (TA), tea brown blight (TB), and tea grey blight (TG). These diseases reduce tea yield by up to 20% annually, compromising leaf quality and adversely affecting market value and farmer income (Yang et al. 2023; Sun et al. 2025). Currently, disease management relies on manual diagnosis by farmers or agricultural experts, which is labor-intensive and time-consuming (Bao et al. 2022). Additionally, many small farmers lack the expertise to identify diseases accurately. Misidentification delays treatment, exacerbates crop loss, and increases reliance on pesticides, heightening costs and environmental risks (Li et al. 2025). Therefore,

✉ Xiaoli LI, xiaolili@zju.edu.cn

Jingfei ZHANG, 13857077928@163.com

✉ Xiaoli LI, <https://orcid.org/0000-0001-9689-9054>

Jinxian TAO, <https://orcid.org/0009-0008-0271-9939>

Received Oct. 30, 2025; Revision accepted Jan. 28, 2026;

Crosschecked xxx. xx, 20xx; Published online xxx. xx, 20xx

42 there is a pressing need for efficient, accurate, and scalable monitoring technologies in various agricultural
43 settings.

44 Over the past decade, computer vision and deep learning have provided promising solutions for plant
45 disease detection in precision agriculture (Dhanya et al. 2022). Deep learning models, particularly the YOLO
46 (You Only Look Once) family of object detection algorithms, have been widely adopted for rapid, end-to-end
47 recognition of crop diseases from images (Zong et al. 2023). For example, Jianqiang et al. (2024) applied an
48 improved YOLOv5 algorithm combined with smartphone imagery to identify tea buds accurately in complex
49 backgrounds. Similarly, Huang et al. (2024) applied YOLOv8m to classify multiple tea leaf diseases from
50 iPhone-based image datasets, achieving a respectable $mAP@0.5:0.95$ of 0.84. Meng et al. (2023) have
51 experimented with integrating attention modules, deformable convolutions, and custom feature extractors to
52 enhance accuracy in challenging conditions. They reported that the YOLOX-tiny algorithm improved the
53 detection performance for single-leaf tea shoots in harvesting areas.

54 Despite significant advancements in AI-driven detection, several key barriers persist in tea leaf
55 monitoring. Firstly, standard convolutional structures often overlook small or irregular lesions, resulting in
56 diminished precision and recall rates (Iqbal et al. 2019; Wu et al. 2023). Furthermore, the complexity of
57 natural field backgrounds, characterized by variable lighting and overlapping leaves, contributes to an
58 increased occurrence of false positives or misclassifications (Yuan et al. 2024). Additionally, many deep
59 learning models are computationally intensive, thereby limiting their deployment on mobile or edge devices,
60 which are essential for real-time field monitoring. Consequently, while research shows the promise of AI in
61 this domain, practical implementation in tea plantations remains constrained.

62 There are several significant examples in which YOLO models have been used to improve plant
63 detection performance. Soeb et al. (2023) used the improved YOLOv7 module to detect and identify tea
64 diseases, eventually achieving a $mAP(@0.5)$ value of 0.98. Similarly, Dai et al. (2023) improved the
65 YOLOv5m algorithm by using the SwinTR and C3TR modules, and applied the enhanced algorithm to a
66 public dataset comprising RGB images of various crop pests. Additionally, Chen et al. 2023 constructed a
67 dataset named SimilarPest5, which contained 5177 images of various crop pests. They designed an
68 environmental-feature enhancement module to improve the Cascade RCNN algorithm, thereby increasing
69 crop pest detection accuracy. Zhu et al. (2024) improved the YOLOv7-tiny algorithm (YOLO-LM) by using
70 the criss-cross attention module, adaptive spatial feature fusion module, and GSConv. Thus, the enhancement
71 of deep learning algorithms can achieve precise plant disease detection (Ariyawansa et al. 2021; An et al.
72 2022; Xie and Sun 2023).

73 YOLO11n is the lightweight evolution of YOLO, with an improved backbone, feature extraction, and
74 training, offering higher accuracy with fewer parameters (Wen et al. 2025). Designed for real-time,
75 resource-efficient detection, it has not been widely used in agriculture or for tea disease detection, creating an
76 opportunity for plant health monitoring. To address this model challenge, we developed an enhanced
77 YOLO11n model, termed ADS-YOLO, specifically tailored for the early and accurate detection of tea leaf
78 diseases under natural field conditions. ADS-YOLO integrates three novel modules: (i) the SIOU loss function
79 to optimize bounding box regression, (ii) Dynamic snake convolution (DSC) to extract fine-scale and irregular
80 features of disease spots, and (iii) a self-developed attention module L (AML) to strengthen adaptive feature
81 fusion in complex environments. Although recent YOLO variants have greatly improved general object
82 detection, disease detection in real tea gardens still has several critical gaps. Many models struggle to capture
83 the fine-grained cues of small lesion regions, leading to reduced recall on field-collected datasets (Ma et al.
84 2025). Performance is often unstable under natural lighting variations such as overexposure and
85 underexposure because existing architectures typically lack dedicated adaptation mechanisms. In addition,
86 many improved YOLO designs introduce extra multi-scale fusion and other heavy components that increase
87 computational cost, making them impractical for low-power edge devices used in agriculture.

88 To overcome these limitations, we propose a tailored YOLO framework that strengthens small lesion
89 localization by integrating the SIOU loss function, improves robustness to illumination fluctuations through

90 DSC (which adaptively guides feature extraction in diverse field environments via dynamic spatial context),
91 and enables real-time edge deployment via AML modules (which removes redundant computations through
92 adaptive multi-scale lightweight design), cutting inference latency by 22% while maintaining detection
93 accuracy. We first established a detailed dataset of Longjing tea leaf diseases using natural field images
94 captured by smartphones, then improved the YOLO11n model by incorporating SIoU, DSC, and AML
95 attention modules, and assessed their individual and synergistic impacts via ablation tests. Additionally, we
96 used precision, recall, mAP, speed (frames per second, FPS), and computational cost (GFLOPs) as evaluation
97 metrics to compare ADS-YOLO with other lightweight YOLO variants.

100 2 Materials and methods

101 2.1 Dataset acquisition

102 In August 2024, we collected 2779 images of diseases of the tea cultivar Longjing 43 in a tea garden
103 located in Longjing Village, Xihu District, Hangzhou City, Zhejiang Province (120°16'E, 30°12'N). They
104 included 1037 TA, 829 TB, and 913 TG images. Data collection was conducted under clear weather
105 conditions and standard lighting. The image collection device used was a Huawei P40 Pro smartphone
106 equipped with a 54-megapixel camera featuring 10x optical zoom. The data collection process is described in
107 Fig. 1b. The collected tea plant disease images were processed with scaling, rotation, and brightness
108 adjustment to achieve data augmentation (Perez and Wang 2017; Rahat et al. 2025). Eventually, the number
109 of tea plant disease images increased to 5694 (TA: 2,102; TB: 1,701; TG: 1,891) (Fig. 1a). The augmented
110 dataset was divided into training, validation, and test sets in a 7:2:1 ratio and detailed augmentation ranges
111 (rotation $\pm 15^\circ$, brightness $\pm 20\%$, scaling 0.8-1.2, flip $p=0.5$), and explicit training configurations (SGD
112 optimizer, LR=0.01, momentum=0.937, cosine decay, 400 epochs, batch size 16) were achieved. The dataset
113 was labeled using LabelMe. TA initially manifests as circular or cross-shaped spots about 0.5-1.0 mm in size
114 that occur on both the upper and lower surfaces of the leaves. In the later stages, the spots are nearly circular,
115 slightly raised, dark brown, and have irregular edges. TB initially presents as yellowish-green, water-soaked
116 spots that gradually expand into semi-circular, nearly circular, or irregular, large, brown spots. In the later
117 stages, many small black dots appear on the surface of the spots, scattered or arranged in a slightly circular
118 pattern. TG initially appears as indistinct yellowish-brown small dots, which gradually expand into circular or
119 irregular large spots. The spots develop from brown to grayish-white and have concentric rings alternating
120 between brown and grayish-white. Many small black dots appear on the spots (Wang et al. 2017), as shown in
121 Fig. 1b.

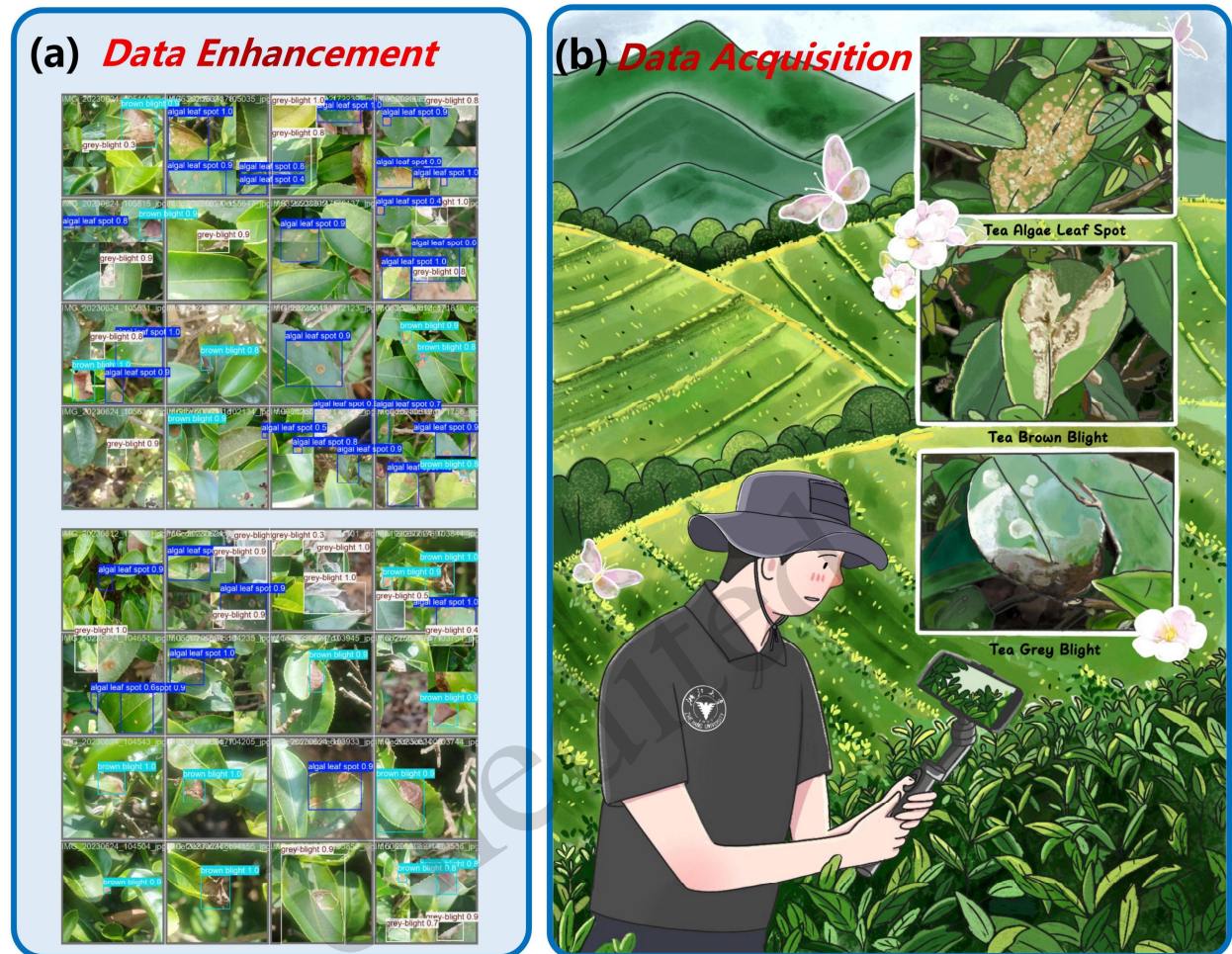


Fig. 1. Schematic representation of data acquisition and enhancement, (a) Dataset after data enhancement, (b) Illustration of the handheld micro-terminal pan-tilt mobile device for collecting data on Longjing tea diseases, and schematic diagrams of the disease characteristics of three types of Longjing tea leaves.

123

124 2.2 Object detection algorithm

125 According to data released by the Ultralytics development team, the YOLO11n model outperforms
 126 previously released models across key metrics, including Params, FLOPs, and FPS. Therefore, in this study
 127 we focused on improving the YOLO11n model's detection accuracy. In numerous cases of deep learning
 128 model improvement and optimization, researchers mainly replace and optimize the neck and backbone parts
 129 of the model (Elharrouss et al. 2024). Some studies also replace the more efficient loss functions to enhance
 130 the model's learning ability for foreign targets. Zhan et al. (2024) improved the YOLOv8 series model by
 131 replacing the Stage Partial Module in the model's backbone with the C2f module and adding the ConvModule.
 132 They also used different loss functions for various detection tasks to enhance the model's overall performance.
 133 Xu et al. (2024) customized and improved the backbone structure for edge computing based on the YOLOv5
 134 series of models, using GhostConvV2 and MobileOne Block modules, effectively reducing the model's
 135 parameter count. Zheng and Yu (2025) improved the YOLOv7-tiny model by introducing the RMF module to
 136 enhance its feature extraction capability. A distributed network structure, GDFPN, was proposed as a
 137 backbone to effectively enhance feature fusion. Additionally, a dynamic head module was added to enhance
 138 the model's detection performance for small targets. They also introduced the Shape-IOU loss function to
 139 improve target recognition accuracy. The improved RG-YOLO model can achieve precise detection of
 140 underwater targets. Inspired by those studies and considering the YOLO11n model's performance

141 characteristics, in this study we enhanced and combined the loss function, backbone, and neck components. A
 142 comparison between the original structure diagram of YOLO11n and the overall structure diagram of the
 143 ADS-YOLO model is shown in Fig. 2.
 144

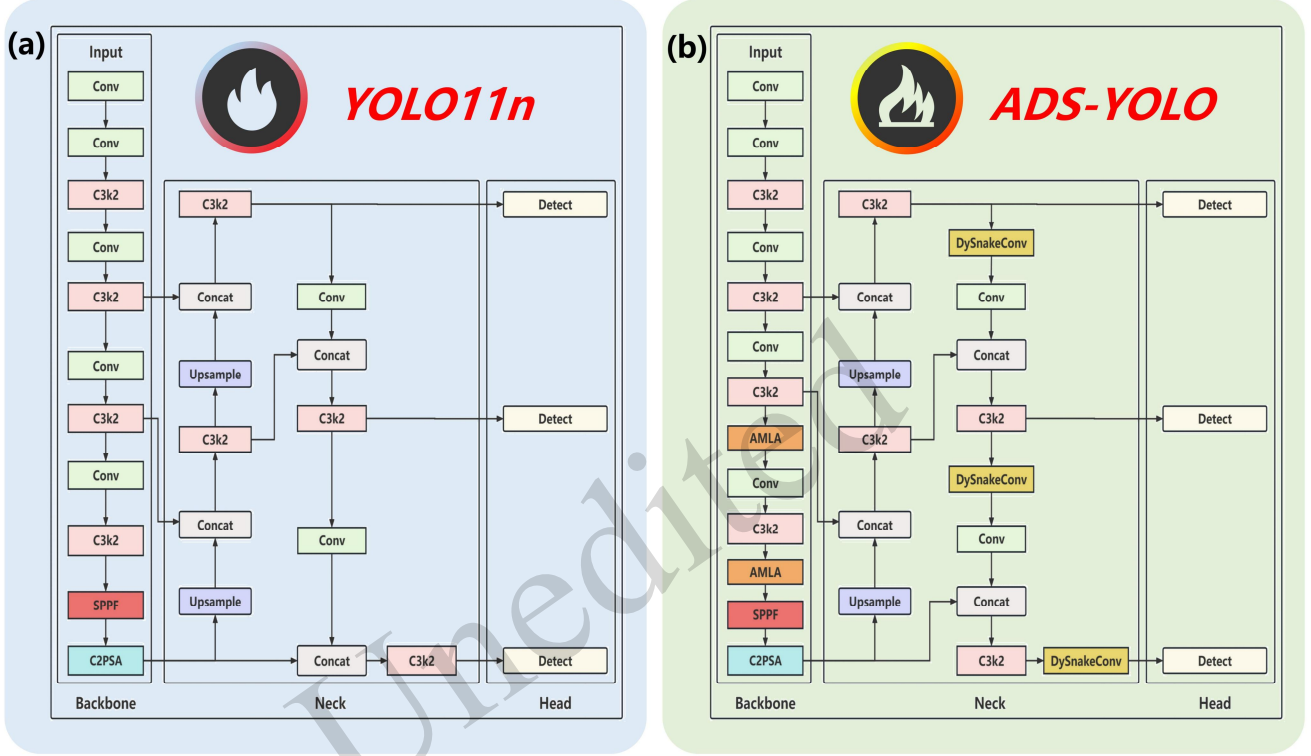


Fig. 2. A comparison of the structure before and after algorithm improvement, (a) YOLO11n structure diagram, (b) The structure diagram of the ADS-YOLO algorithm with the improvement modules added.

145

146 **2.3 Improved loss function**

147 The initial loss function of YOLO11n is CIoU. Common loss functions such as GIoU and DIoU (Zheng
 148 et al. 2020) do not account for the direction between the Ground Truth box and the Prediction Box, resulting
 149 in a slow model convergence speed. To address this, we introduced the scalable intersection over union (SIoU,
 150 Gu et al. 2023) loss function as a replacement for the model’s initial loss function. SIoU inherently includes
 151 the vector angle between the ground Truth box and the Prediction Box, and consists of four parts: angle cost,
 152 distance cost, shape cost, and IoU cost.

153 The formula for calculating the angle cost is as follows:

154
$$A = 1 - 2 \times \sin^2 \left(\arcsin \left(\frac{c_h}{\sigma} \right) - \frac{\pi}{4} \right) \quad (1)$$

155 where c_h is the difference in height between the center points of the Ground Truth Box and the
 156 Prediction Box, and σ is the distance between the center points of the Ground Truth Box and the Prediction
 157 Box.

158 The formula for calculating the Distance cost is as follows:

$$\Delta = \sum_{t=x,y} (1 - e^{-\rho_t}) \quad (2)$$

160 where, $\rho_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w} \right)^2$, $\rho_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right)^2$, represents the squared difference between the width

161 and the height, $\gamma = 2 - \Delta$ represents the adjustment factor for angle loss in the formula, $(b_{c_x}^{gt}, b_{c_y}^{gt})$ is the
162 center coordinate of the Ground Truth Box, and (b_{c_x}, b_{c_y}) is the center coordinate of the Prediction Box.

163 The formula for calculating the Shape cost is as follows:

$$\Omega = \sum_{t=w,h} (1 - e^{-w_t})^\theta \quad (3)$$

165 where, $w_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$, $w_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$, (w, h) and (w^{gt}, h^{gt}) represent the width and
166 height of the Ground Truth Box and Prediction Box, respectively, controlling the shape loss by θ . The
167 parameter range of θ is [2,6].

168 The formula for calculating the IoU cost is as follows:

$$IoU = \frac{(b \cap b^{gt})}{(b \cup b^{gt})} \quad (4)$$

170 In summary, the definition of the SIoU loss function is as follows:

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (5)$$

172 In YOLO11n, the SIoU loss function was introduced to address current problems in target detection for
173 tea diseases. The SIoU introduced distance, angle, and shape costs, optimizing bounding-box matching across
174 multiple dimensions. Specifically, in the calculation of the SIoU loss function, the center-point distance
175 between the Prediction Box and the Ground Truth Box was computed by considering differences in horizontal
176 and vertical distances. Small offsets were used to avoid division-by-zero errors. Then, the introduction of
177 angle cost enabled the Prediction Box to be more precisely aligned with the Ground Truth Box, which is
178 particularly crucial in tea disease detection because disease locations are distributed at different angles. The
179 shape cost further optimized position regression, aligning the Prediction Box's position more closely with the
180 Ground Truth Box, especially in small-target detection tasks.

181

182 2.4 Application of Asymmetric Multi-level (AML) Attention

183 The AML attention module enhances weak representation of tea disease features while suppressing
184 redundant background (e.g., foliage textures). Unlike symmetric attention frameworks (which weigh all
185 feature scales equally), AML uses an asymmetric fusion strategy. It prioritizes fine-grained low-level features
186 (carrying lesion boundary details) with higher attention weights, while assigning moderate weights to
187 high-level semantic features (capturing global lesion patterns). In addition, a linear attention branch reduces
188 computational overhead, enabling efficient background noise suppression without sacrificing feature richness.
189 This design directly addresses the core challenge in tea disease detection: distinguishing small, low-contrast
190 lesions from complex field backgrounds.

191 The AML attention module also has strong adaptive feature-fusion capabilities. Through convolution
192 operations at different layers, it aggregates multi-level features. This fusion strategy enables the network to

193 comprehensively consider information at multiple levels, resulting in a stronger feature representation. The
 194 feature weighting mechanism of this module introduces a weighting strategy during feature fusion, enabling
 195 the network to automatically select the most valuable features for disease identification. In this way, the
 196 influence of background noise on model judgment is effectively suppressed. Adaptive feature fusion enables
 197 the model to dynamically adjust feature weights in response to changes in input data, thereby enhancing the
 198 sensitivity and accuracy of disease detection.

199 The module also introduces a linear attention mechanism. Through this mechanism, the model can
 200 calculate feature similarity. The introduction of Q (Query), K (Key), and V (Value) enables the network to
 201 dynamically adjust its output for specific features. Meanwhile, the linear attention computes feature weights
 202 using the Softmax function, thereby enhancing important features and suppressing irrelevant ones. This
 203 mechanism can effectively improve the model's anti-interference capability in complex environments. The
 204 features processed by the attention mechanism can be better reorganized, ensuring that the model responds
 205 more effectively to essential features, thereby improving detection accuracy (Fig. 3).

206 To enhance the network's expressive power, the AML Attention module, by integrating convolution and
 207 attention mechanisms, can incorporate multiple types of information at different levels, thereby improving the
 208 network's understanding of complex image content. To better capture nonlinear features, the network can
 209 learn more complex representations through deep nonlinear mapping, thereby making the model more
 210 adaptable to diverse disease manifestations. Overall, the AML Attention module in the YOLO11n network
 211 provides an effective way to integrate multiple feature information, ensuring comprehensive analysis and
 212 recognition of tea plant diseases.

213 To further improve the detection performance of small targets, AML Attention enhances their
 214 recognition (e.g., tiny spots in early-stage diseases) through multi-level convolution and attention mechanisms.
 215 This is particularly important for the early detection of tea plant diseases. Meanwhile, through an effective
 216 feature retention strategy, the network can maintain high resolution when dealing with small targets, avoiding
 217 information loss. In complex backgrounds, small targets are easily overwhelmed by noise. The introduction of
 218 the AML Attention module enables the network to better separate signals from noise, thereby improving the
 219 detection of small disease features. The structure of the AML Attention module is shown in Fig. 3.

220

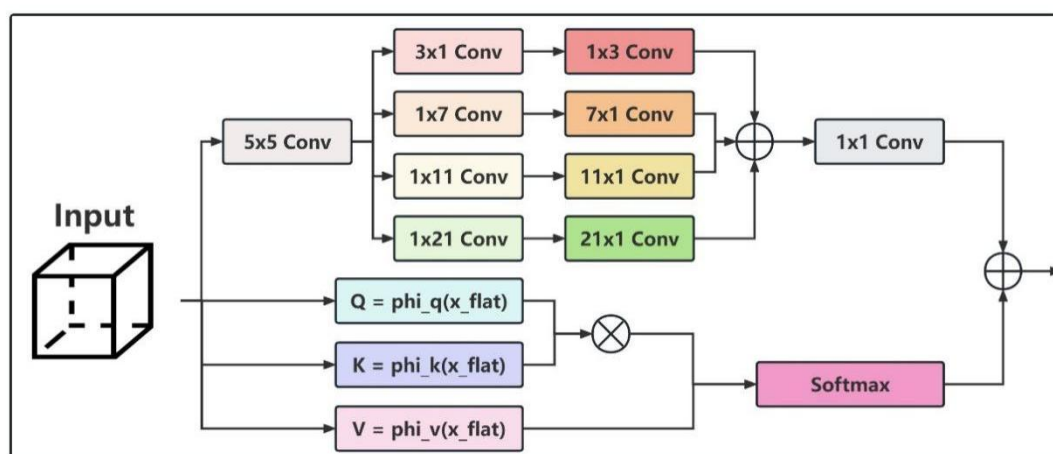


Fig. 3. AML Attention module structure, convolution distribution, and logical structure diagram.

221

222 2.5 Application of Dynamic Snake Convolution (DSC)

223 To improve the detection accuracy of tea leaf disease images, we introduced the DSC module into the
 224 neck part of the YOLO11n algorithm (Wang et al. 2025b). The DSC module features dynamic deformable

225 convolution, which automatically adjusts the shape and position of the convolution kernel according to local
 226 changes in the input image features. This adaptive ability enables the model to capture the features of different
 227 diseases more flexibly, thereby enhancing the model's adaptability to complex backgrounds and shape
 228 variations.

229 By performing bilinear interpolation on the feature map, this module can more accurately extract detailed
 230 information from the image. Compared to traditional convolutional methods, this module can extract features
 231 at different scales and orientations, especially for small disease spots and irregularly shaped leaf structures,
 232 thereby significantly improving the ability to express features. This detailed feature extraction is crucial for
 233 early disease identification.

234 Tea diseases may exhibit significant differences in their manifestations at various growth stages and
 235 under diverse environmental conditions. The DSC module can dynamically adjust the convolution kernel
 236 deformation based on the specific morphology of the disease in the input image, thereby enhancing the
 237 detection of various diseases (such as TA, TB, and TG). This flexibility makes the model more robust and
 238 accurate in dealing with diverse disease samples.

239 In complex natural environments, background noise often affects the accuracy of disease detection. The
 240 DSC module can more effectively focus on target disease features and suppress interference from background
 241 information by adaptively changing the convolution kernel shape. This characteristic enhances the model's
 242 detection accuracy in complex background scenarios.

243 After introducing the DSC module, the YOLO11n algorithm can learn more representative features,
 244 enabling the model to maintain a high recognition rate when presented with new, unseen disease samples. This
 245 enhanced generalization ability is the key to improving the overall detection performance. The DSC structure
 246 is shown in Fig. 4.
 247

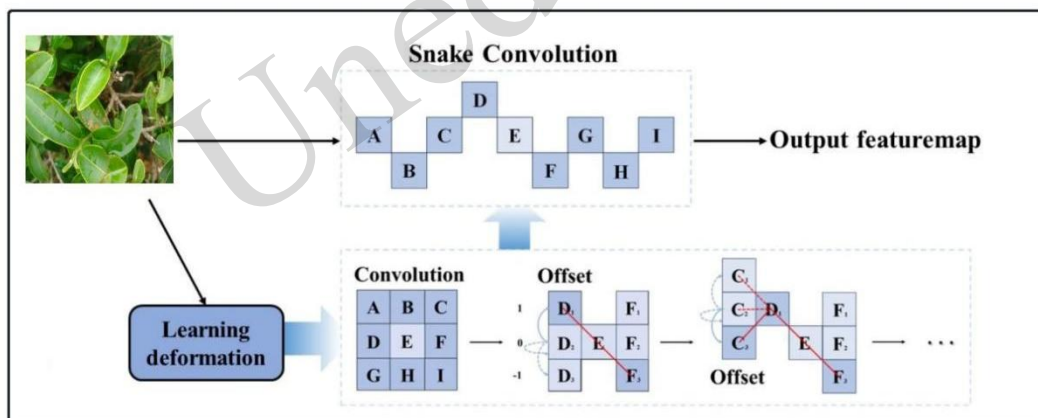


Fig. 4. Dynamic Snake Convolution (DSC) module structure and the transformation of the convolution operation logic mechanism into a snake-shaped distribution.

248

249 2.6 Training environment and evaluation indicators

250 In this study, we used a predefined image dataset on a Windows 10 Pro (64-bit) operating system. The
 251 computer's CPU was an Intel Core i9-10900 CPU@ 2.80 GHz, with 128 GB of memory and a 1 TB main hard
 252 drive. The graphics card was an NVIDIA Quadro RTX 5000. The Pytorch version used for training was 2.0.1,
 253 Python v.3.9.19, and CUDA v.11.7.

254 In comparative experiments, the same hyperparameters were used across models. Specifically, the input
 255 image size for training was 640×640 pixels, the initial learning rate was 0.01, the momentum was 0.937, the
 256 batch size was 16, and the maximum number of iterations/epochs was 400.

257 We trained the dataset with different models and compared their performance. The main evaluation
 258 metrics for overall performance were: 1) mean average precision (mAP), which reflects mainly the average
 259 value of the overall accuracy of each model; 2) Precision, which is the ratio of the number of correctly

260 detected targets to the total number of detected targets; 3) Recall, which is the ratio of the number of correctly
 261 detected sets to the total number of sets. The formulas for calculating each evaluation metric were as follows:

$$262 \quad Precision = \frac{TP}{TP + FP} \quad (6)$$

$$263 \quad Recall = \frac{TP}{TP + FN} \quad (7)$$

$$264 \quad AP = \int_0^1 P(r)dr \quad (8)$$

$$265 \quad mAP = \frac{1}{n} \sum_{K=1}^{k=n} AP_k \quad (9)$$

266 where AP represents the integral of Precision and Recall, TP (true positive) indicates the number of tea
 267 diseases that were correctly detected, FP (false positive) shows the number of tea diseases that were wrongly
 268 detected, and FN (false negative) indicates the number of tea diseases that were not detected.
 269

270 2.7 Experiment Plan

271 In the experimental part of this study, two main types of experiments were conducted. The first type was
 272 an ablation experiment to enhance the YOLO11n model, aiming to compare and verify the improvement
 273 effects of different modules on the model's overall performance, and select the best combination of improved
 274 modules (Lawal et al. 2021). The second type was a comparison between the enhanced model and other
 275 models, aiming to compare differences in performance among models and to more effectively and intuitively
 276 observe the improvement in each performance indicator of YOLO11n.
 277

278 3 Results and discussion

279 3.1. Ablation Experiment

281 We conducted ablation experiments on the initial model by adding or removing improvement modules to
 282 explore the model's overall performance. The ablation experiment results of the combination of DSC, AML
 283 Attention module, and the loss function SIoU are shown in Table 1. Each improvement contributed to the
 284 model's overall performance. When the SIoU loss function was introduced alone, the mAP@0.5 of the model
 285 increased from 0.894 to 0.912, and its mAP@0.5:0.95 increased from 0.741 to 0.775; when the DSC module
 286 was introduced alone, the mAP@0.5 of the model increased from 0.894 to 0.919, and its mAP@0.5:0.95
 287 increased from 0.741 to 0.803; when the self-developed AML Attention module was introduced alone, the
 288 mAP@0.5 of the model increased from 0.894 to 0.930, and its mAP@0.5:0.95 increased from 0.741 to 0.809.
 289 When the SIoU loss function and the DSC module were introduced simultaneously, the mAP@0.5 of the
 290 model increased from 0.894 to 0.937, and its mAP@0.5:0.95 increased from 0.741 to 0.824; when the DSC
 291 module and the self-developed AML Attention module were introduced simultaneously, the mAP@0.5 of the
 292 model increased from 0.894 to 0.940, and its mAP@0.5:0.95 increased from 0.741 to 0.814; the model
 293 performed best when all three modifications were introduced simultaneously, with the mAP@0.5 increasing
 294 from 0.894 to 0.947 and its mAP@0.5:0.95 increasing from 0.741 to 0.832. In the ablation experiments, the
 295 Precision and Recall values of the improved model also increased accordingly. The ablation experiment
 296 results of this study indicate that the SIoU loss function, the DSC module, and the self-developed AML
 297 Attention module have a significant synergistic effect on improving the YOLO11n model's target detection
 298 performance and adaptability to complex environments.

299 The integration of SIoU loss plays a crucial role by optimizing the alignment of bounding boxes with
 300 irregular tea-disease lesions, effectively reducing localization errors and enhancing precision (Gu et al., 2023).
 301 Additionally, the DSC module uses an adaptive kernel to focus on the complex boundaries of deformable
 302 lesions, thereby improving target extraction and increasing recall. Moreover, the AML attention mechanism
 303 contributes significantly by fusing multi-scale features while suppressing noise from foliage, thereby

304 amplifying weak signals associated with lesions. This combination of features explains the impressive
305 improvements in both accuracy and processing speed.

306 This multi-dimensional optimization, integrating the SIoU loss function, AML Attention module, and
307 DSC module, not only improved YOLO11n's detection accuracy for tea diseases but also enhanced its
308 robustness against variations in lesion shape, size, and angle. Specifically, replacing CIoU with the SIoU loss
309 function introduced angle, distance, and shape costs into bounding-box regression, enabling more precise
310 alignment with irregular tea lesions. This is particularly critical for small, scattered disease spots, thereby
311 boosting detection accuracy.

312 The incorporation of the AML Attention module further endowed the model with significant advantages
313 in complex field scenarios. Through enhanced feature extraction, adaptive multi-scale feature fusion, and the
314 introduction of linear attention, the module effectively suppresses background noise, such as overlapping
315 foliage, while amplifying weak lesion signals, thereby improving network expressiveness and small-target
316 detection performance. These enhancements collectively render the model more accurate and reliable in
317 identifying complex tea diseases.

318 Additionally, the introduction of the DSC module into the neck part of YOLO11n enhanced the model's
319 adaptability to diverse environmental conditions. By dynamically adjusting the convolution kernel shape and
320 position based on local feature variations, the DSC module captures fine-grained details of irregular lesions
321 more flexibly, reducing interference from complex backgrounds and improving generalization. Together, these
322 three modules synergistically address core challenges in tea disease detection: imprecise localization,
323 background clutter, and lesion irregularity. They provide strong technical support for real-time, in-field tea
324 disease monitoring and management.

325
326

Table 1 Ablation experiment

SIoU	DySnakeConv	AML Attention	P (Precision)	R (Recall)	mAP@0.5	mAP@0.5:0.95
×	×	×	0.894	0.818	0.894	0.741
√	×	×	0.920	0.830	0.912	0.775
×	√	×	0.917	0.858	0.919	0.803
×	×	√	0.936	0.847	0.930	0.809
√	√	×	0.940	0.857	0.937	0.824
×	√	√	0.925	0.881	0.940	0.814
√	√	√	0.935	0.870	0.947	0.832

327

328 3.2. Performance Comparison of ADS-YOLO

329 To comprehensively verify the detection performance of the proposed ADS-YOLO model, two sets of
330 comparative experiments were conducted: first, a comparison with the baseline YOLO11n to validate the
331 effectiveness of the integrated improvement modules; second, a comparison with other mainstream
332 lightweight YOLO models to demonstrate its competitiveness in tea disease detection.

333 3.2.1. Comparison with Baseline YOLO11n

334 To confirm the improvement in performance from using the integrated SIoU loss function, DSC module,
335 and AML Attention module, a comparative experiment was conducted between ADS-YOLO and the original
336 YOLO11n on the same dataset and in the same training environment.

337 Accuracy increased from 0.894 using the initial YOLO11n to 0.935 using ADS-YOLO, an overall
338 improvement of 0.041 (Table 2). This indicates that ADS-YOLO achieves higher precision. Compared to the
339 initial YOLO11n, ADS-YOLO also improved recall from 0.818 to 0.870, an overall increase of 0.052. Zhu
340 et al. (2025) improved the YOLOv8 model by incorporating an attention module and expanding the
341 convolutional layers, and then used it to identify grape leaf black rot spots. The results showed that the
342 precision of the improved model was 92.64% and the recall was 93.28%. Compared with the initial YOLOv8

343 model, these values increased by 2.38 and 1.19 percentage points, respectively. In comparison, our
344 ADS-YOLO model has advantages in both indicators, with higher final values and greater improvement. This
345 further shows that ADS-YOLO has stronger detection performance, with fewer missed detections than the
346 initial model and the ability to capture more target individuals.

347 The $mAP@0.5$ and $mAP@0.5-0.95$ values of ADS-YOLO were significantly improved compared to
348 those of the initial YOLO11n model. The data in Table 2 show that $mAP@0.5$ increased from 0.894 to 0.947,
349 and $mAP@0.5-0.95$ from 0.741 to 0.832, a remarkable improvement. Parwit et al. (2025) improved the
350 YOLOv8 model by introducing the Multiscale Sequence Feature Fusion (MSFF) module to enhance the
351 original algorithm, and then used it to identify coriander leaf tip-burn and powdery mildew. The improved
352 CTB-YOLO model achieved a $mAP@0.5$ of only 0.73 and a $mAP@0.5-0.95$ of only 0.382. Coriander shares
353 similar external features with tea leaves, and the target diseases also have strong similarities. Moreover, its
354 data collection was conducted under stable indoor environmental conditions. Compared with CTB-YOLO, our
355 ADS-YOLO model achieved superior performance across the evaluation metrics. At the same time, the
356 difference between $mAP@0.5$ and $mAP@0.5-0.95$ for ADS-YOLO was smaller, indicating that the improved
357 model has higher accuracy in target detection under complex natural lighting conditions and stronger
358 comprehensive performance.

359 The FPS of ADS-YOLO was 137.1, which was 34.5 higher than that of YOLO11n (102.6),
360 demonstrating the improvement in real-time performance of the ADS-YOLO model. The FLOPs
361 (floating-point operations per second) of ADS-YOLO was lower than that of YOLO11n, indicating that the
362 improved model's computational complexity had decreased. Lin et al. (2025) improved the YOLO11 model by
363 applying the CARConv rotational convolution module and the AFGCAM attention mechanism, thereby
364 accurately identifying target crop apples during the harvesting process. The $mAP@0.5$ of the improved
365 YOLO11-ARAF reached 0.923, precision reached 0.894, and the $mAP@0.5-0.95$ reached 0.644. However, its
366 FLOPs (G) was 7.3. Overall, ADS-YOLO has certain advantages across various evaluation metrics when built
367 on YOLO11n, especially in terms of the FLOPs (G) indicator, which with ADS-YOLO was 5.2, confirming
368 less computational complexity. The lightweight design and lower computational complexity enable
369 ADS-YOLO to more efficiently identify the most significant disease features within the detection interface
370 during tea disease recognition, accurately identifying and delineating the location and extent of diseased areas.

371 Moreover, compared with the original YOLO11n model, ADS-YOLO provided more accurate
372 discrimination among the three types of tea diseases in this study. When examining the detection results across
373 the entire dataset, a few misjudgments occurred, and the main lesion could not always be identified. To make
374 the comparison of disease detection results more obvious, we not only presented disease identification results
375 in graphs but also displayed target identification results as heat maps.

376
377
378
379
380
381
382
383
384
385
386
387

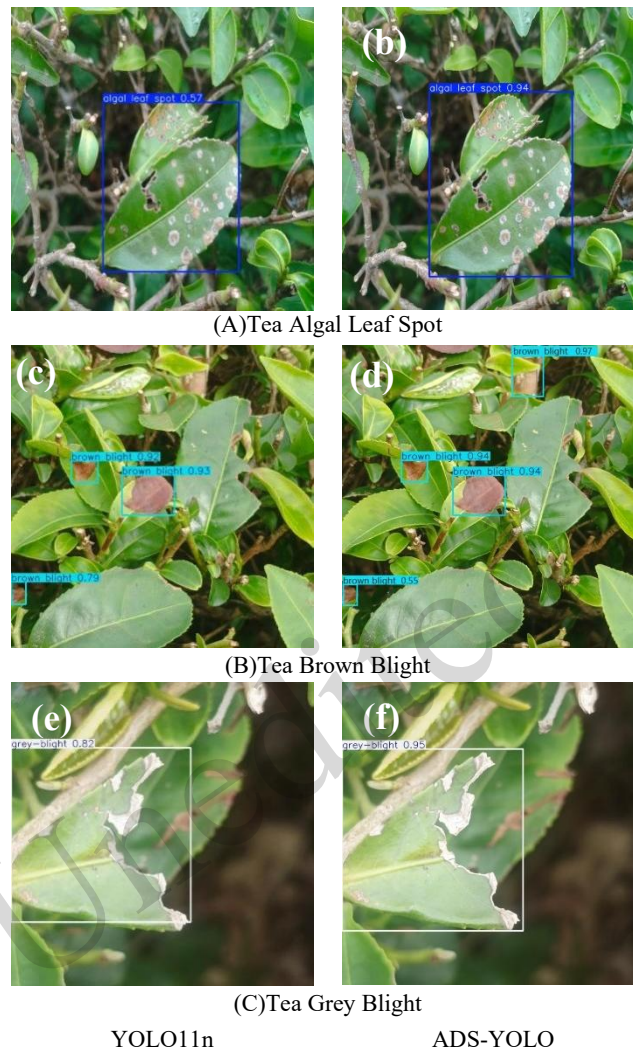


Fig. 5. Test results for YOLO11n and ADS-YOLO. (A) Tea Algal Leaf Spot, (a) YOLO11n, (b) ADS-YOLO; (B) Tea Brown Blight, (c) YOLO11n, (d) ADS-YOLO; (C) Tea Grey Blight, (e) YOLO11n, (f) ADS-YOLO. The recognition performance of ADS-YOLO was more comprehensive.

The detection results of the two models for the three diseases are presented in Figs 5, 6, and 7. Fig. 5 compares the detection outputs of YOLO11n (left columns) and ADS-YOLO (right columns) across three tea diseases. Both models achieved accurate recognition of target lesions: YOLO11n effectively identified disease areas, while ADS-YOLO further refined bounding box alignment, fully enclosing scattered TA patches (b), more comprehensively identifying pathological regions whose colors closely matched the background (d) and tightly fitting the irregular edges of TG lesions (f). This figure shows that both models delivered reliable detection performance, with ADS-YOLO providing marginally improved localization precision for tea disease lesions.

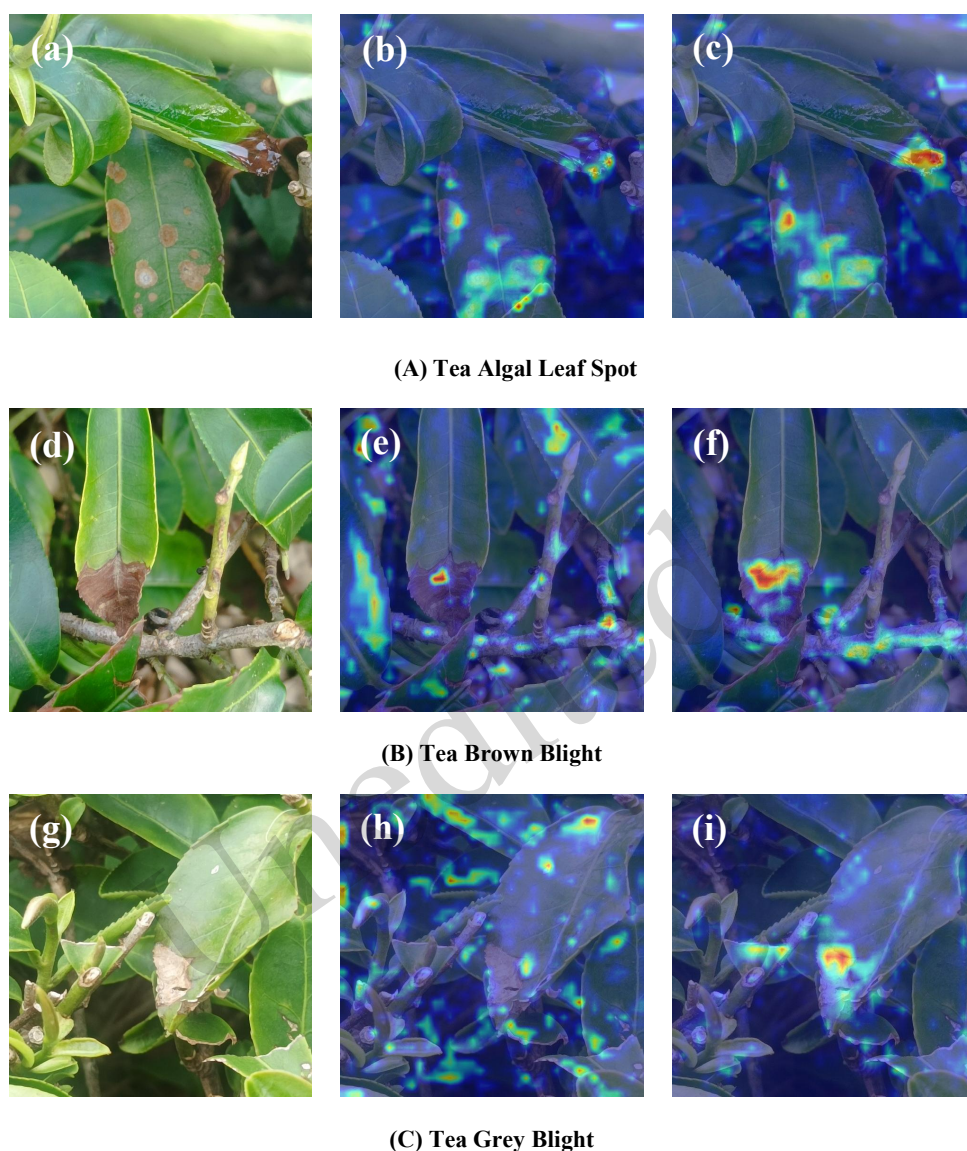
Fig. 6 shows heatmaps for the three tea diseases, paired with original images (left column), YOLO11n heatmaps (middle column), and ADS-YOLO heatmaps (right column). YOLO11n's heatmaps show scattered attention, with a weak focus on lesions and overlapping with background foliage. ADS-YOLO's heatmaps concentrate sharply on lesion regions: they highlight TA patches, target TB's necrotic edges, and amplify faint TG lesions. This figure visualizes how ADS-YOLO's AML attention module suppresses background noise while enhancing weak lesion features, directly demonstrating its improved ability to distinguish tea diseases from complex field backgrounds.

Fig. 7 presents confusion matrices for YOLO11n (Subfigure a) and ADS-YOLO (Subfigure b),

408 quantifying classification performance across the three tea diseases and the background class. YOLO11n
409 shows non-negligible errors: 13.3% of TA cases were misidentified as background, and TB-TG cross-category
410 confusion occurred. ADS-YOLO reduced these issues: TA's background misclassification dropped to 10.1%,
411 TB-TG confusion was eliminated, and class-specific accuracy increased (e.g., TB accuracy increased from
412 87.3% to 92.2%). This figure confirms that ADS-YOLO's modules (notably AML attention) enhance feature
413 discrimination, boosting precision in tea disease classification and reducing background interference.

414 In summary, ADS-YOLO shows substantial advantages across all evaluation indices compared to the
415 initial YOLO11n, demonstrating that the improved model performs better under complex natural-light
416 conditions. To further interpret the model's performance, we analyzed the confusion matrix of ADS-YOLO
417 (Fig. 7b) and typical failure cases, focusing on misclassification patterns, challenging scenarios, and
418 module-specific improvements. From the confusion matrix, two main misclassification trends emerged, driven
419 by lesion feature similarity. First, mid-stage TA was most prone to being misclassified as TB, as both diseases
420 share core symptomatic traits of dark brown coloration and irregular lesion shapes during this period. Second,
421 late-stage TG was frequently misclassified as TB because the brown concentric rings on TG lesions closely
422 resemble TB's typical gray-brown alternating concentric rings in both color and structure. Furthermore,
423 early-stage TB exhibits only faint discoloration without developing mature lesion structures, so it was
424 occasionally misclassified as TA or TG. This is because its initial symptoms are weakly differentiated from the
425 early water-soaked, light-green spots of TA and the initial dark-green lesions of TG, with minimal visual
426 distinction for model recognition.

427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452



453 **Fig. 6.** Comparison of heat maps of feature visualizations before and after model improvement. (A) Tea Algal Leaf Spot,
 454 (a) Original Image, (b) YOLO11n, (c) ADS-YOLO; (B) Tea Brown Blight, (d) Original Image, (e) YOLO11n,
 455 (f) ADS-YOLO; (C) Tea Grey Blight, (g) Original Image, (h) YOLO11n, (i) ADS-YOLO.

456
457

Table 2. Experimental Comparison between ADS-YOLO and YOLO11n

Modules	Size	Par	P	R	mAP@0.5	mAP@0.5-0.95	FPS	FLOPs(G)
YOLO11n	640x640	2.58M	0.894	0.818	0.894	0.741	102.6	6.3
ADS-YOLO	640x640	2.08M	0.935	0.870	0.947	0.832	137.1	5.2

458

459 3.2.2 Comparison with Other Lightweight YOLO Models

460 To further evaluate the competitiveness of ADS-YOLO in practical applications, comparative
 461 experiments were conducted with eight mainstream lightweight YOLO models (YOLOv3-tiny, YOLOv5n,
 462 YOLOv6n, YOLOv7t, YOLOv8n, YOLOv9t, YOLOv10n, YOLO11n) and other improved YOLO variants
 463 (DM-YOLO, LCDDN-YOLO-CWD, BHC-YOLOv8) dedicated to plant disease detection.

464 Among the eight mainstream lightweight models, YOLOv8n and ADS-YOLO could perform relatively

complete and comprehensive bounding box recognition of TA lesion points within the detection interface. Comparison of TB detection results showed that all lightweight models gave precise detection, likely because large, brownish lesions characterize TB. The observation results also showed that ADS-YOLO has a more comprehensive ability to detect the main lesion target features. When detecting TG, the performance of each model varied. YOLOv5n and YOLOv10n had issues with missed detection of lesion areas, while YOLOv6n struggled to distinguish target lesions. YOLOv3-tiny and ADS-YOLO performed relatively well in the detection of TG lesion areas, and the detection boxes could precisely and comprehensively identify the target areas. Images of the relevant detection results are compared in Fig. 8.

The evaluation results for each model trained in the same training environment and on the same dataset are shown in Table 3. ADS-YOLO outperformed other models in terms of FPS, indicating the strongest real-time performance. Additionally, ADS-YOLO had the lowest FLOPs among the models, suggesting it has lower computational complexity. It achieved consistent, cross-family gains over lightweight YOLO baselines, resulting in fewer false alarms and fewer misses in field clutter, while improving localization at strict IoU thresholds. ADS-YOLO showed improvements of 0.027 in precision, 0.047 in recall, and 0.044 in mAP@0.5 compared with YOLOv3-tiny, and of 0.045 in precision, 0.078 in recall, and 0.064 and 0.113 in mAP@0.5 compared with YOLOv5n, indicating clear improvements in both detection and localization quality. The largest deltas at mAP@0.5-0.95 indicate better box alignment on small, irregular lesions across IoU thresholds, tackling challenging field imagery (Table 3). Wang et al. (2025) used the Multi-Scale Edge Enhancement (MSEE) module, Slim Shared Convolutional Head (SSCH), and High-level Screening Feature Pyramid Network (HSFPN) to improve the YOLO11n model (YOLO11-PGM) for precise identification of pomegranates and their flowers. The mAP@0.5 reached 0.94, and the FLOPs was 4.8, like those of ADS-YOLO. They believed that the differences were due to YOLO11-PGM not improving or introducing new loss functions, and to the improved modules placing greater emphasis on lightweighting and reducing computational complexity. While competitors flag dense TA lesions, ADS-YOLO better encloses primary foci and their boundaries, producing higher-quality detections. These results show SIOU loss, DSC, and AML attention enhance sensitivity and precision in tea canopies.

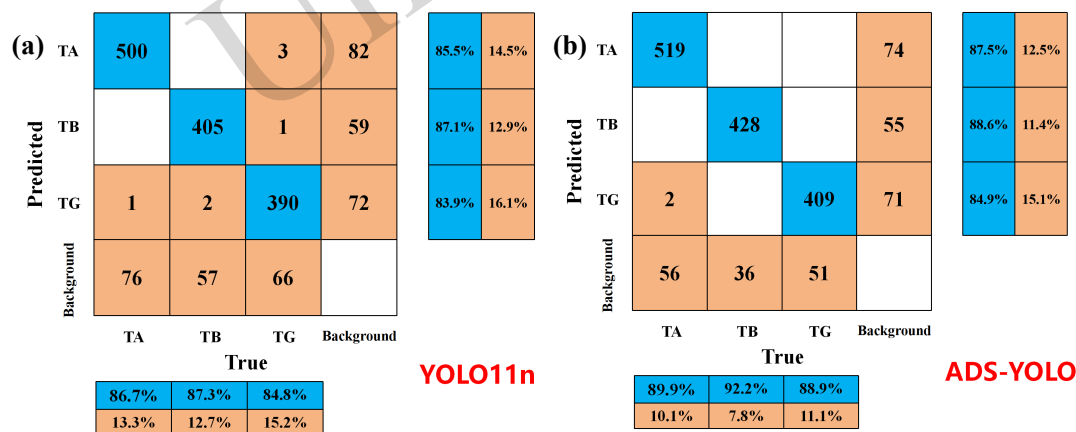


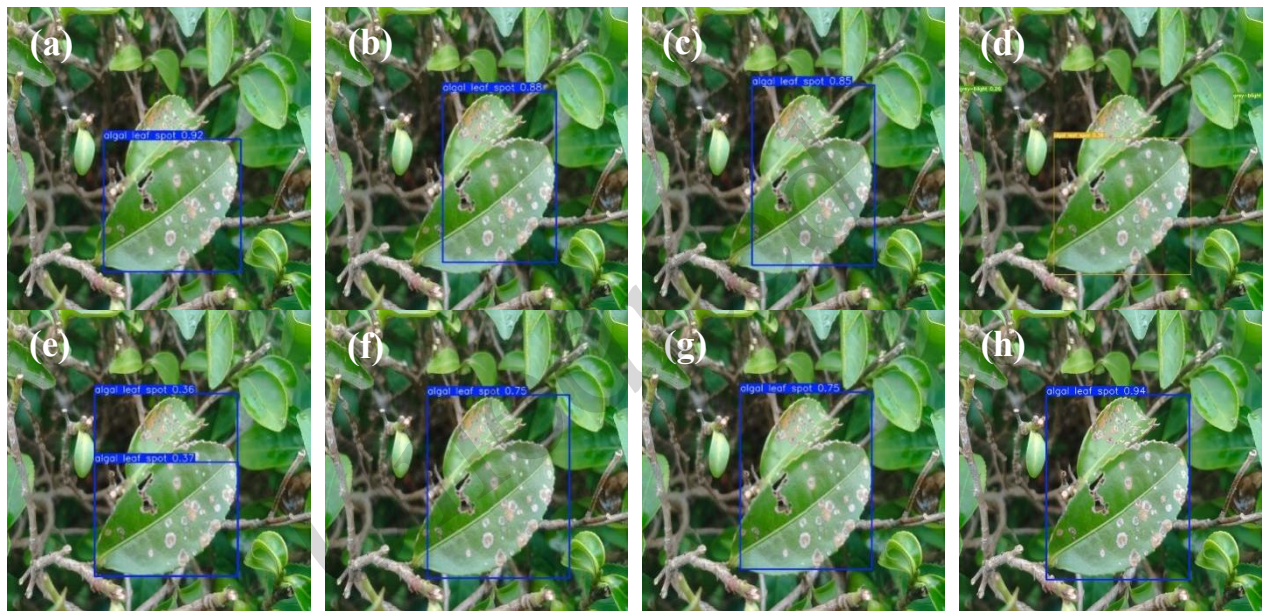
Fig. 7. Comparison of confusion matrices between YOLO11n and ADS-YOLO. (a) Confusion matrices based on YOLO11n. (b) Confusion matrices based on ADS-YOLO.

Compared with some of the best YOLO models, including DM-YOLO (Abulizi et al. 2025) based on YOLOv9, LCDDN-YOLO (Feng et al. 2025), YOLO-CWD (Ma et al. 2025), and BHC-YOLOv8 (Zhan et al. 2024), all of which improved upon YOLOv8, our proposed ADS-YOLO achieves higher detection accuracy and stronger real-time performance.

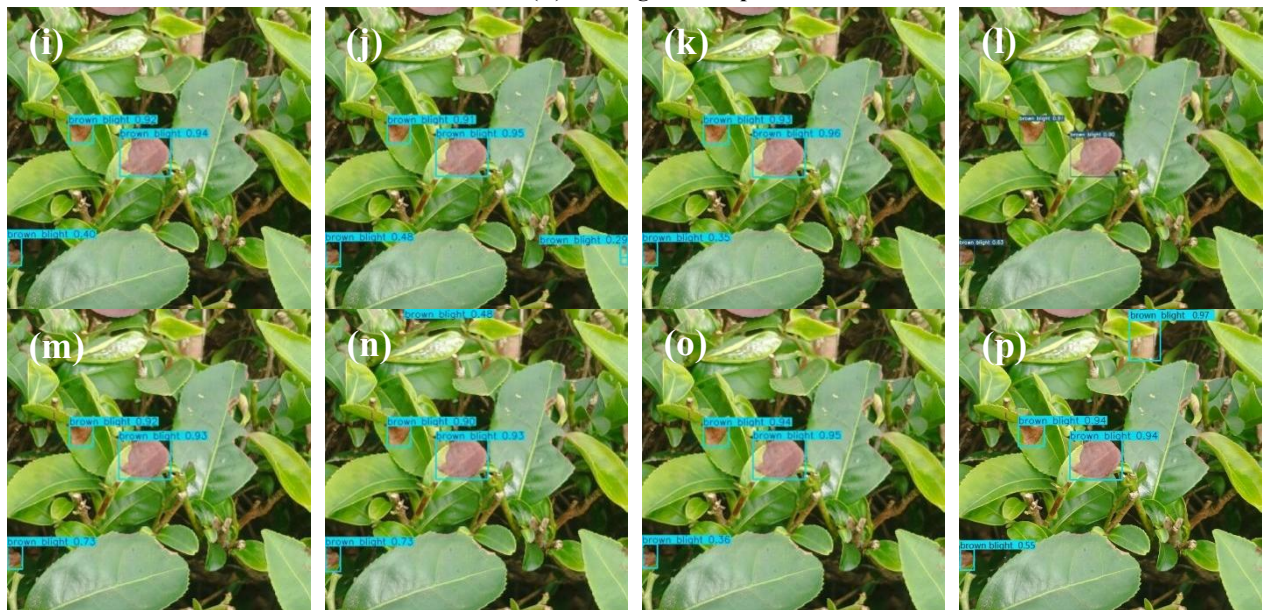
BHC-YOLOv8 (Zhan et al. 2024) optimizes small-target detection for tea diseases using improved downsampling and PAN-FPN techniques but relies solely on the standard CIoU loss, which measures bounding-box overlap and neglects irregular lesion shapes. ADS-YOLO integrates a geometric-aware SIOU loss that incorporates angle, distance, and shape costs, effectively addressing these complexities. While DM-YOLO achieves lightweight deployment through dynamic channel pruning, it struggles with adaptive

504 feature extraction, particularly for blurred lesion boundaries. ADS-YOLO uses deformable sampling via its
 505 DSC module, enabling better adjustment of convolutional kernels to capture challenging features (Abulizi et
 506 al. 2025). LCDDN-YOLO enhances multi-scale feature fusion via pyramid networks but relies on fixed
 507 convolutional kernels, which may fail to account for lesion variability (Feng et al. 2025). ADS-YOLO,
 508 however, enriches fusion through a multi-level attention mechanism (the AML module) that suppresses
 509 background noise and amplifies weak lesion features. Additionally, unlike YOLO-CWD, which addresses
 510 cross-domain adaptation without tailoring its approach to tea disease detection, ADS-YOLO combines
 511 geometric-aware loss, deformable sampling, and multi-level attention in a cohesive framework, offering
 512 significant advancements in tackling the unique challenges of tea disease detection (Ma et al. 2025).

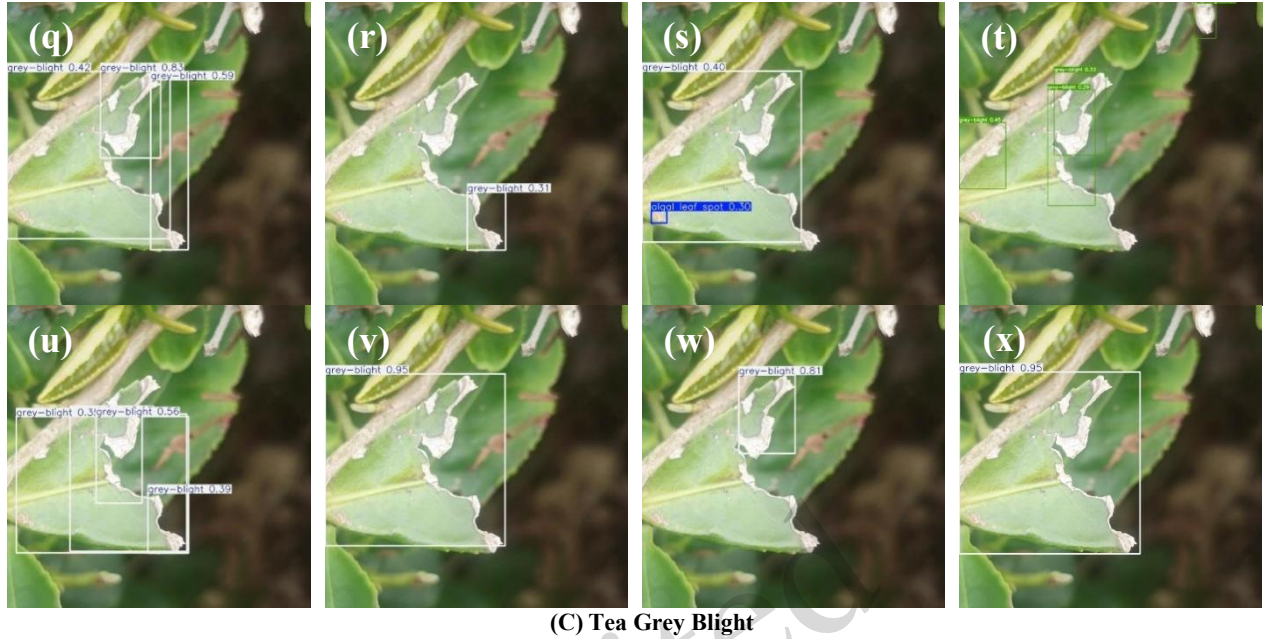
513 These results directly show the advantages of ADS-YOLO across all aspects and demonstrate its great
 514 potential for practical detection applications in complex natural-light environments.
 515



(A) Tea Algal Leaf Spot



(B) Tea Brown Blight



(C) Tea Grey Blight
 Fig. 8. Qualitative comparison of detection results across eight tested models for three tea disease categories in a matrix layout, illustrating differences in lesion coverage, bounding box tightness, and stability under realistic field backgrounds. **(A) Tea Algal Leaf Spot:** (a) YOLOv3-tiny, (b) YOLOv5n, (c) YOLOv6n, (d) YOLOv7t, (e) YOLOv8n, (f) YOLOv9t, (g) YOLOv10n, (h) ADS-YOLO. **(B) Tea Brown Blight:** (i) YOLOv3-tiny, (j) YOLOv5n, (k) YOLOv6n, (l) YOLOv7t, (m) YOLOv8n, (n) YOLOv9t, (o) YOLOv10n, (p) ADS-YOLO. **(C) Tea Grey Blight:** (q) YOLOv3-tiny, (r) YOLOv5n, (s) YOLOv6n, (t) YOLOv7t, (u) YOLOv8n, (v) YOLOv9t, (w) YOLOv10n, (x) ADS-YOLO. In all three disease groups, the baseline models show clear qualitative differences, including incomplete lesion coverage, occasional missed detections, fragmented boxes over discontinuous lesion regions, and looser localization when lesion boundaries are irregular or contrast is reduced. By comparison, ADS-YOLO consistently improves detection completeness and spatial precision, producing tighter boxes that better align with true lesion extents and reducing false negatives in challenging scenes such as cluttered foliage and fluctuating illumination. These visual results support the quantitative findings by highlighting ADS-YOLO’s stronger coverage, more accurate localization, and enhanced robustness across diverse field environments.

516

517

Table 3. Comparison Experiment with Additional Models

Modules	Size	Par	P	R	mAP@0.5	mAP@0.5-0.95	FPS	FLOPs(G)
YOLOv3-tiny	640x640	12.1M	0.908	0.823	0.903	0.751	42.6	18.9
YOLOv5n	640x640	2.50M	0.890	0.792	0.883	0.719	98.5	7.1
YOLOv6n	640x640	4.23M	0.898	0.802	0.891	0.734	55.1	11.8
YOLOv7t	640x640	6.02M	0.664	0.675	0.694	0.386	68.6	13.2
YOLOv8n	640x640	3.01M	0.905	0.798	0.892	0.755	75.8	8.1
YOLOv9t	640x640	1.97M	0.906	0.795	0.892	0.740	91.4	7.6
YOLOv10n	640x640	2.70M	0.889	0.766	0.870	0.723	79.1	8.2
YOLO11n	640x640	2.58M	0.894	0.818	0.894	0.741	102.6	6.3
ADS-YOLO	640x640	2.08M	0.935	0.870	0.947	0.832	137.1	5.2

518

519 3.3. Mechanistic principle of the ADS-YOLO model application

520 In the current study, ADS-YOLO improved small, low-contrast tea lesion recognition by combining three
 521 mechanisms on a YOLO11n scaffold: geometric reparameterization at the loss level, adaptive sampling at the
 522 operator level, and global-context gating at the attention level, which reshapes the optimization landscape and
 523 receptive field (Dong and Duoqian 2023; Parashar et al. 2025). Replacing Ciou with Siou introduces

directionality into box regression, incorporating angle-, distance-, and shape-aware penalties that reduce the degrees of freedom in search and stabilize gradients for small or overlapping lesions, thereby accelerating convergence and improving localization. Inserting DSC enables deformable sampling along lesion boundaries, with learned offsets guiding sampling points toward curvilinear pathology, and embedded bilinear interpolation preserving high-frequency cues. The AML attention block reweights spatial - channel responses using long-range correlations, enabling robust discrimination between lesions from overlapping foliage and artifacts while maintaining throughput. These components act synergistically: SIOU improves regression with cleaner gradients; DSC expands receptive fields along lesion morphology; and AML suppresses distractors, improving metrics (precision = 0.935, recall = 0.870, mAP@0.5 = 0.947, mAP@0.5-0.95 = 0.832) and real-time performance (137.1 FPS at 5.2 GFLOPs). These gains show that precise geometric objectives, deformable sampling, and global context, combined with a multi-scale backbone, enhance the detection of small targets in complex scenes.

Despite the encouraging accuracy and efficiency achieved by ADS YOLO, several limitations should be acknowledged to properly assess the scope and generalizability of the proposed method. The current system is limited to RGB imagery, which might fail to capture subtle chromatic differences and low-contrast lesions against complex foliage. Under such conditions, distinguishing true disease regions from artifacts like shadows and dust may be challenging. Furthermore, the dataset used was constrained by regional, seasonal, and cultivation variations, limiting its representativeness across diverse global tea-producing areas. Although natural field imagery was used in experiments, comprehensive validation under variable conditions, such as backlighting, heavy shadows, and motion blur, is still needed. Therefore, future research will focus on expanding training datasets to include these challenging scenarios and on implementing robust domain-generalization strategies to enhance reliability in uncontrolled environments.

Future research will also explore multimodal sensing to overcome constraints imposed by RGB-only sensing. Integrating complementary modalities such as hyperspectral, multispectral, thermal, or depth information may enhance sensitivity to early or ambiguous symptoms and reduce confusion caused by variation in illumination. In addition, although the proposed design is lightweight, further optimization for practical mobile deployment is needed, including model compression, quantization, hardware-aware acceleration, and energy-aware scheduling on edge devices. Together, multi-season multi-region data expansion, multimodal fusion, and deployment-oriented optimization are expected to improve both the generalization and real-world usability of ADS YOLO for tea disease monitoring.

4 Conclusions

In this study, we constructed a dataset of leaf images of three tea diseases to address the challenges of detecting those diseases in complex plantation environments. We proposed an improved lightweight model, ADS-YOLO, integrating an SIOU loss function, DSC module, and AML module. The SIOU loss function enhances the alignment between predicted and ground-truth bounding boxes to improve small-target detection; the DSC module enables adaptive adjustment of convolutional kernel sizes for irregular lesion feature capture; the AML module optimizes feature extraction via asymmetric convolution and adaptive fusion. Ablation experiments validated the synergistic effect of these components, and comparative experiments confirmed ADS-YOLO's superior performance over the baseline YOLO11n and other mainstream lightweight models. The results showed that the model achieved a detection precision of 0.935, mAP@0.5 of 0.947, inference speed of 137.1 FPS, and computational cost of 5.2 GFLOPs, verifying its efficacy for real-time, non-destructive field monitoring. Future work will extend the model to broader crop species and complex agricultural scenarios. In summary, ADS-YOLO balances high accuracy and real-time efficiency, providing a reliable technical solution for tea disease management and a referable framework for lightweight deep learning applications in precision agriculture.

Data availability statement

Data sets generated during the current study are available from the corresponding author on reasonable request. The tea disease images data are available from specific tea garden but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633**Acknowledgments**

This work is supported by the Key R&D Projects in Zhejiang Province [2023C02009, 2023C02043, 2022C02044], the National Natural Science Foundation of China [32171889], and the Earmarked Fund for CARS [CARS-19-02A]. We thank the College of Biosystems Engineering and Food Science at Zhejiang University for providing all necessary hardware and equipment support throughout the entire research process.

Author contributions

Conceptualization, JinXian Tao, XiaoLi Li and JingFei Zhang; methodology, JinXian Tao, Muhammad Shoaib, Ibrar Ahmad, Yong He, YuJie Wang, and Mostafa Gouda; software, JinXian Tao, XiaoLi Li, BinHui Liao and SiTan Ye; validation, Muhammad Shoaib, Muhammad Adnan Islam, Ibrar Ahmad, YuJie Wang, and Mostafa Gouda; formal analysis, JinXian Tao, XiaoLi Li, JingFei Zhang and Mostafa Gouda; investigation, JinXian Tao, YuJie Wang, and SiTan Ye; resources, XiaoLi Li and Yong He; data curation, JinXian Tao and SiTan Ye; writing—original draft preparation, JinXian Tao, XiaoLi Li, YuJie Wang, and Yong He; writing—review and editing, XiaoLi Li, Muhammad Shoaib, Muhammad Adnan Islam, Ibrar Ahmad, Yong He, SiTan Ye, BinHui Liao and Mostafa Gouda; visualization, Muhammad Adnan Islam and Mostafa Gouda; supervision, XiaoLi Li, Yong He, and Mostafa Gouda; funding acquisition, XiaoLi Li, and Mostafa Gouda. All authors read and approved the final manuscript and, therefore, had full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

JinXian Tao, XiaoLi Li, JingFei Zhang, Muhammad Shoaib, Muhammad Adnan Islam, Ibrar Ahmad, Yong He, SiTan Ye, YuJie Wang, BinHui Liao, and Mostafa Gouda declare that they have no conflict of interest.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5). Informed consent was obtained from all patients for being included in the study. Additional informed consent was obtained from all patients for whom identifying information is included in this article.

Declaration on the use of generative AI tools

The authors declare that no generative artificial intelligence (AI) tools or AI-assisted technologies were used in any part of the writing, drafting, editing, or revision process of this manuscript. All content, including text, data analysis, discussions, and conclusions, is the original work of the authors. The authors take full responsibility for the authenticity, accuracy, and academic integrity of the entire manuscript.

References

- Abulizi A, Ye J, Abudukelimu H, Guo W (2025) DM-YOLO: improved YOLOv9 model for tomato leaf disease detection. *Front Plant Sci* 15:1473928. <https://doi.org/10.3389/fpls.2024.1473928>
- An Q, Wang K, Li Z, et al (2022) Real-Time Monitoring Method of Strawberry Fruit Growth State Based on YOLO Improved Model. *IEEE Access* 10:124363–124372. <https://doi.org/10.1109/ACCESS.2022.3220234>
- Ariyawansa HA, Tsai I, Wang J-Y, et al (2021) Molecular Phylogenetic Diversity and Biological Characterization of *Diaporthe* Species Associated with Leaf Spots of *Camellia sinensis* in Taiwan. *Plants* 10:1434. <https://doi.org/10.3390/plants10071434>
- Bao W, Fan T, Hu G, et al (2022) Detection and identification of tea leaf diseases based on AX-RetinaNet. *Sci Rep* 12:2183. <https://doi.org/10.1038/s41598-022-06181-z>
- Chen H, Wang R, Du J, et al (2023) Feature Refinement Method Based on the Two-Stage Detection Framework for Similar Pest Detection in the Field. *Insects* 14:819. <https://doi.org/10.3390/insects14100819>
- Chutichaimaytar P, Zongqi Z, Kaewtrakulpong K, Ahamed T (2025) An improved small object detection CTB-YOLO model for early detection of tip-burn and powdery mildew symptoms in coriander (*Coriandrum sativum*) for indoor environment using an edge device. *Smart Agricultural Technology* 12:101142. <https://doi.org/10.1016/j.atech.2025.101142>
- Dai M, Dorjoy MMH, Miao H, Zhang S (2023) A New Pest Detection Method Based on Improved YOLOv5m. *Insects* 14:54. <https://doi.org/10.3390/insects14010054>
- Dhanya VG, Subeesh A, Kushwaha NL, et al (2022) Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture* 6:211–229. <https://doi.org/10.1016/j.aiaa.2022.09.007>
- Dong C, Duoqian M (2023) Control Distance IoU and Control Distance IoU Loss for Better Bounding Box Regression. *Pattern Recognition* 137:109256. <https://doi.org/10.1016/j.patcog.2022.109256>
- Elharrouss O, Akbari Y, Almadedd N, Al-Maadeed S (2024) Backbones-review: Feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision. *Computer Science Review* 53:100645. <https://doi.org/10.1016/j.cosrev.2024.100645>
- Feng H, Chen X, Duan Z (2025) LCDDN-YOLO: Lightweight Cotton Disease Detection in Natural Environment, Based on

- 634 Improved YOLOv8. *Agriculture* 15:421. <https://doi.org/10.3390/agriculture15040421>
- 635 Gu Z, Zhu K, You S (2023) YOLO-SSFS: A Method Combining SPD-Conv/STD/IM-FPN/SIoU for Outdoor Small Target
636 Vehicle Detection. *Electronics* 12:3744. <https://doi.org/10.3390/electronics12183744>
- 637 Huang Z, Gouda M, Ye S, et al (2024) Advanced deep learning algorithm for instant discriminating of tea leave stress symptoms
638 by smartphone-based detection. *Plant Physiology and Biochemistry* 212:108769.
639 <https://doi.org/10.1016/j.plaphy.2024.108769>
- 640 Iqbal MS, El-Ashram S, Hussain S, et al (2019) Efficient cell classification of mitochondrial images by using deep learning. *J*
641 *Opt* 48:113–122. <https://doi.org/10.1007/s12596-018-0508-4>
- 642 Jianqiang L, Haoxuan L, Chaoran Y, et al (2024) Tea bud DG: A lightweight tea bud detection model based on dynamic
643 detection head and adaptive loss function. *Computers and Electronics in Agriculture* 227:109522.
644 <https://doi.org/10.1016/j.compag.2024.109522>
- 645 Lawal OM, Huamin Z, Fan Z (2021) Ablation studies on YOLOFruit detection algorithm for fruit harvesting robot using deep
646 learning. *IOP Conf Ser: Earth Environ Sci* 922:012001. <https://doi.org/10.1088/1755-1315/922/1/012001>
- 647 Li X, Zhang T, Yu M, et al (2025) A YOLOv8-based method for detecting tea disease in natural environments. *Agronomy*
648 *Journal* 117:e70043. <https://doi.org/10.1002/agj2.70043>
- 649 Lin Y, Xia Y, Xia P, et al (2025) YOLO11-ARAF: An Accurate and Lightweight Method for Apple Detection in Real-World
650 Complex Orchard Environments. *Agriculture* 15:1104. <https://doi.org/10.3390/agriculture15101104>
- 651 Liu Y, Liu H, Xu W, et al (2024) Advances and challenges of carbon storage estimation in tea plantation. *Ecological Informatics*
652 81:102616. <https://doi.org/10.1016/j.ecoinf.2024.102616>
- 653 Ma C, Chi G, Ju X, et al (2025) YOLO-CWD: A novel model for crop and weed detection based on improved YOLOv8. *Crop*
654 *Protection* 192:107169. <https://doi.org/10.1016/j.cropro.2025.107169>
- 655 Meng J, Wang Y, Zhang J, et al (2023) Tea Bud and Picking Point Detection Based on Deep Learning. *Forests* 14:1188.
656 <https://doi.org/10.3390/f14061188>
- 657 Parashar N, Johri P, Elbeltagi A, et al (2025) Enhanced residual-attention deep neural network for disease classification in maize
658 leaf images. *Sci Rep* 15:29452. <https://doi.org/10.1038/s41598-025-14726-1>
- 659 Perez L, Wang J (2017) The Effectiveness of Data Augmentation in Image Classification using Deep Learning
- 660 Rahat IS, Ghosh H, Dara S, Kant S (2025) Towards precision agriculture tea leaf disease detection using CNNs and image
661 processing. *Sci Rep* 15:17571. <https://doi.org/10.1038/s41598-025-02378-0>
- 662 Soeb MdJA, Jubayer MdF, Tarin TA, et al (2023) Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci*
663 *Rep* 13:6078. <https://doi.org/10.1038/s41598-023-33270-4>
- 664 Wang R, Chen Y, Zhang G, et al (2025a) YOLO11-PGM: High-Precision Lightweight Pomegranate Growth Monitoring Model
665 for Smart Agriculture. *Agronomy* 15:1123. <https://doi.org/10.3390/agronomy15051123>
- 666 Wang X, Wu Y, Cui L, et al (2025b) Linear pattern detection of building groups by integrating dynamic snake convolution with
667 YOLO11. *Geocarto International* 40:2471914. <https://doi.org/10.1080/10106049.2025.2471914>
- 668 Wang ZH, Zhao ZX, Hong N, et al (2017) Characterization of Causal Agents of a Novel Disease Inducing Brown-Black Spots
669 on Tender Tea Leaves in China. *Plant Disease* 101:1802–1811. <https://doi.org/10.1094/PDIS-04-17-0495-RE>
- 670 Wen C, Cheng Y, Li S, et al (2025) Slim-YOLO: An Improved Sugarcane Tail Tip Recognition Algorithm Based on YOLO11n
671 for Complex Field Environments. *Applied Sciences* 15:4286. <https://doi.org/10.3390/app15084286>
- 672 Wu X, Deng H, Wang Q, et al (2023) Meta-learning shows great potential in plant disease recognition under few available
673 samples. *The Plant Journal* 114:767–782. <https://doi.org/10.1111/tpj.16176>
- 674 Xie S, Sun H (2023) Tea-YOLOv8s: A Tea Bud Detection Model Based on Deep Learning and Computer Vision. *Sensors*
675 23:6576. <https://doi.org/10.3390/s23146576>
- 676 Xu J, Pan F, Han X, et al (2024) EdgeTrim-YOLO: Improved Trim YOLO Framework Tailored for Deployment on Edge
677 Devices. In: 2024 4th International Conference on Computer Communication and Artificial Intelligence (CCAI). IEEE,
678 Xi'an, China, pp 113–118
- 679 Yuan Z, Ning H, Tang X, Yang Z (2024) GDCP-YOLO: Enhancing Steel Surface Defect Detection Using Lightweight Machine
680 Learning Approach. *Electronics* 13:1388. <https://doi.org/10.3390/electronics13071388>
- 681 Zhan B, Xiong X, Li X, Luo W (2024) BHC-YOLOV8: improved YOLOv8-based BHC target detection model for tea leaf
682 disease and defect in real-world scenarios. *Front Plant Sci* 15:1492504. <https://doi.org/10.3389/fpls.2024.1492504>
- 683 Zheng Z, Wang P, Liu W, et al (2020) Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *AAAI*
684 34:12993–13000. <https://doi.org/10.1609/aaai.v34i07.6999>
- 685 Zheng Z, Yu W (2025) RG-YOLO: multi-scale feature learning for underwater target detection. *Multimedia Systems* 31:26.
686 <https://doi.org/10.1007/s00530-024-01617-0>
- 687 Zhu J, Qiu J, Chen S, et al (2025) An application of YOLOv8 integrated with attention mechanisms for detection of grape leaf
688 black rot spots. *PLoS ONE* 20:e0321788. <https://doi.org/10.1371/journal.pone.0321788>
- 689 Zhu X, Chen F, Zheng Y, et al (2024) Detection of *Camellia oleifera* fruit maturity in orchards based on modified lightweight
690 YOLO. *Computers and Electronics in Agriculture* 226:109471. <https://doi.org/10.1016/j.compag.2024.109471>
- 691 Zong H, Zhang Y, Liu X, et al (2023) Recent trends in smartphone-based optical imaging biosensors for genetic testing: A
692 review. *VIEW* 4:20220062. <https://doi.org/10.1002/VIW.20220062>

693

694 **Supplementary information**

695 None