International Journal of Advanced Engineering and Management Research

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Teabud-Yolo11++: An Annotation-Efficient Semi-Supervised Framework for Machine Vision-Based Tea-Bud Detection in Intelligent Tea Harvesting.

Md Morshed Alam[1], Jiangsheng Gui[1]*
[1]School of Computer Science and Technology, Zhejiang Sci-Tech University
Hangzhou, Zhejiang 310018, China.

doi.org/10.51505/ijaemr.2026.11601 URL: http://dx.doi.org/10.51505/ijaemr.2026.11601

Received: Oct 21, 2026 Accepted: Nov 02, 2026 Online Published: Nov 14, 2026

Abstract
Accurate tea-bud detection is essential for intelligent tea harvesting, yield estimation, and precision tea production. However, practical deployment remains challenging because of dense bud distributions, severe occlusion, variable illumination, and the high cost of large-scale manual annotation. To address these challenges, we propose TeaBud-YOLO11++, an annotation-efficient semi-supervised framework that extends YOLO11 through teacher–student learning and a novel Density-Aware Confidence Recalibration (DACR) mechanism that adaptively recalibrates pseudo-label confidence according to local bud density. By improving pseudo-label reliability in densely populated scenes, the proposed framework enables more effective use of unlabeled data while reducing annotation dependency. Experiments on a field dataset of 1,500 images containing 68,895 manually annotated tea buds showed that TeaBud-YOLO11++ achieved 93.33% precision, 91.21% recall, 95.08% mAP@0.5, and 75.45% mAP@0.5:0.95, outperforming representative YOLO11 variants, RT-DETR-L, and Faster R-CNN. Furthermore, DACR reduced the Expected Calibration Error by 28.8%, Furthermore, DACR reduced the Expected Calibration Error by 28.8%, while environmental robustness, mixed-domain evaluation, and repeated-training analyses provided additional evidence of the framework's effectiveness under diverse plantation conditions. These results demonstrate that TeaBud-YOLO11++ provides a robust, annotation-efficient machine vision solution for intelligent tea harvesting and precision tea production.

Keywords: Agricultural automation, Intelligent tea harvesting, Object detection, Precision agriculture, Semi-supervised learning, Tea bud detection.

1. Introduction

Tea is one of the most economically important plantation crops worldwide, and the livelihood of millions of people in many tea-producing nations depends on it either directly or indirectly through provision of rural livelihood opportunities, agricultural employment, and foreign exchange earnings. The commercial value and quality of tea products largely depend on

Vol. 11, No. 06; 2026

ISSN: 2456-3676

harvesting tender shoots at the appropriate developmental stage (Aaqil et al., 2023; Wang, Xi, et al., 2025). Typically, only the apical bud together with two to three young leaves is harvested for the production of high-quality tea. Tea harvesting remains a labor-intensive operation and represents one of the largest production costs in tea cultivation (Abhiram, 2024; Huang et al., 2025). Sustained tea production has been increasingly challenging to maintain due to labor shortages and increasing wages, as well as a decline in the number of skilled workers recently, which underscores the pressing need for smart harvesting technology that can increase harvesting efficiency and decrease labor dependence (Abeyrathna et al., 2026; Fahmida Islam et al., 2022; Singh et al., 2020).

Recent advances in precision agriculture have accelerated the adoption of machine vision and artificial intelligence for automating agricultural operations (Soleimani et al., 2026). Machine vision-based object detection has become a powerful method for monitoring crops, estimating yield, diagnosing diseases, harvesting fruit, and managing fields intelligently. Among them, precise tea-bud detection is considered the most basic requirement for automated tea harvesting, since harvesting robots depend on accurately locating the tender buds prior to conducting picking operations. Nevertheless, tea bud detection in practice still faces challenges such as the miniature size of tea buds, high spatial density, severe occlusion, complex backgrounds, and illumination variations in field environments. These factors degrade the accuracy of detection significantly and limit the outdoor application of existing vision-based harvesting techniques (Hou et al., 2023; Wang, Li, et al., 2025; Yu et al., 2024; Zhang et al., 2023).

Deep learning-based object detectors, particularly the YOLO family, have been widely adopted in agricultural machine vision because of their high detection accuracy and real-time performance (Badgujar et al., 2024; Pugazhendi et al., 2026). Several studies have successfully applied YOLO-based detectors for tea-bud detection with promising results. Despite these advances, most current methods are based on fully supervised learning, which necessitates a large amount of manually annotated images. Annotating densely populated tea-field images is labor-intensive, time-consuming, and costly because numerous small and partially occluded tea buds must be accurately labeled. Consequently, the high annotation cost remains a major barrier to the large-scale adoption of intelligent tea-harvesting systems (Yang et al., 2024; Zhang et al., 2025).

In addition to annotation dependency, several practical challenges remain insufficiently addressed. Most existing work focuses on classic detection accuracy and provides little assessment of model robustness when tested in actual field environments with low illumination, motion blur, and various weather conditions. In addition, there has been a lack of studies focusing on cross-domain generalization among different tea plantations, as various tea plantations are very different from each other due to factors including geographical location, cultivation techniques, imaging tools, and weather conditions. These domain shifts bring about much performance loss when the models learn from one plantation. The trustworthiness of prediction confidence has scarcely been studied, although confidence estimation is critical for

Vol. 11, No. 06; 2026

ISSN: 2456-3676

reliable decision-making in automated agricultural systems, and a false prediction may result in an immediate impact on the quality of harvest and efficiency of operation (Hernández Hernández et al., 2025; Sbai, 2025).

These limitations hinder the practical deployment of existing tea-bud detection systems in commercial tea plantations (Li et al., 2025; Zhang et al., 2026). A successful intelligent harvesting scheme should consider not just high detection accuracy but also low annotation cost, strong robustness, reliable prediction confidence, and superior cross-domain generalization among tea plantations. To the best of our knowledge, existing studies have not comprehensively addressed these practical requirements within a unified framework.

To address these challenges, this study proposes TeaBud-YOLO11++, an annotation-efficient semi-supervised framework for machine vision-based tea-bud detection in intelligent tea harvesting. The proposed framework integrates teacher-student semi-supervised learning to effectively utilize unlabeled field images while reducing manual annotation requirements. Furthermore, a novel Density-Aware Confidence Recalibration (DACR) module is introduced to improve prediction confidence in densely populated tea-bud scenes. Unlike previous studies that primarily emphasize detection accuracy, this work provides a mixed-domain evaluation of annotation efficiency, robustness under challenging field conditions, confidence calibration, cross-domain generalization, and deployment potential, thereby enhancing the practical applicability of intelligent tea harvesting systems.

The main contributions of this study are summarized as follows:

1. A YOLO11-based semi-supervised tea-bud detection network is designed to make full use of unlabeled field images which can greatly alleviate manual annotation cost and achieve a high detection accuracy for intelligent tea harvesting.
2. In this article, we propose a novel DACR (Density-Aware Confidence Recalibration) module to enhance the reliability of prediction confidence and detection stability in extremely crowded tea plantation areas.
3. The framework is evaluated under diverse environmental conditions, with additional analyses of robustness, confidence calibration, and geographical domain shift. These evaluations provide evidence of its potential applicability to intelligent tea harvesting.

This study contributes toward robust, scalable, and annotation-efficient machine vision systems for intelligent tea harvesting and next-generation precision agriculture.

The motivation and overall workflow of the proposed framework are summarized in figure 1. Unlike conventional tea-bud detection approaches that primarily emphasize detection accuracy, the proposed system considers a deployment-oriented evaluation protocol that includes semi-supervised learning performance, confidence calibration based on DACR, environmental robustness testing, geographical robustness assessment, and annotation efficiency measures.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

These complementary factors bring the system for detecting tea buds to a more practically reliable and scalable level for the development of intelligent and autonomous tea harvesting under real plantation circumstances.

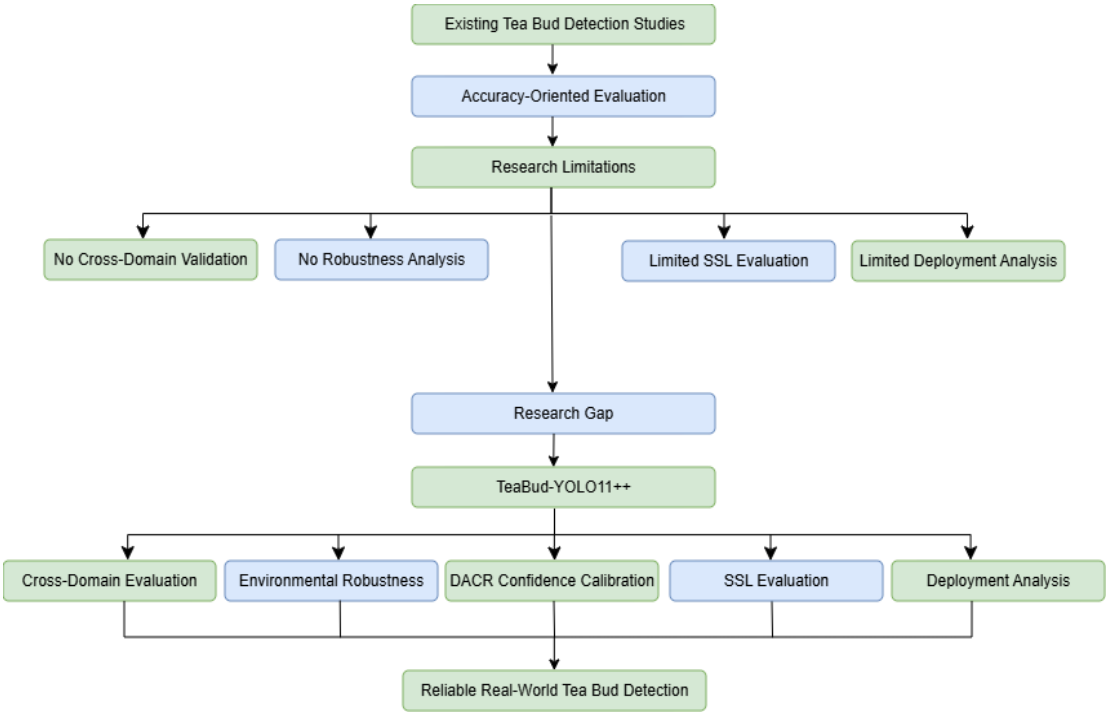


Figure 1. Research Motivation and Conceptual Framework of TeaBud-YOLO11++.

In addition to the detection accuracy improvement, the purpose of the proposed framework is to enable low annotation cost, reliable detection, and precise tea production for practical intelligent tea harvesting under various plantation environments.

2. Related Work

2.1 Deep Learning-Based Tea Bud Detection

The developments in deep learning and computer vision have facilitated the integration of intelligent technology into precision agriculture, enabling the intelligent detection of crops, diseases, yields, and fruits, as well as providing harvest assistance. Accurate and timely identification of harvestable tea buds is essential for improving harvesting efficiency, maintaining product quality, reducing labor requirements, and enhancing plantation productivity. Vision-based detection systems can support these objectives by identifying suitable harvesting targets under field conditions. With the labor shortages and increasing production cost posing a major challenge to tea growing around the world, dependable vision-based harvesting systems have gained more attention as a potential solution to enhance the sustainability and productivity of tea production (Yang et al., 2024; Yang et al., 2023; Yildiz et al., 2026).

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Traditional tea-bud detection methods are mostly based on handcrafted features and combined with traditional machine-learning classifiers (Wang et al., 2026; Wang et al., 2024). Although traditional methods demonstrated the feasibility of tea-bud identification, their reliance on handcrafted features made them sensitive to illumination changes, complex backgrounds, canopy variations, and occlusion. As a result, their applicability in practice was not large-scale in commercial tea production. The introduction of convolutional neural networks (CNNs) and deep learning detectors significantly improved feature representation, localization accuracy, and robustness to challenging field environments, ultimately paving the way for intelligent tea harvesting. Among advanced deep learning-based object-detection methodologies, the YOLO family is regarded as the most efficient for agricultural vision with high detection accuracy, computational efficiency, and capability of real-time inference.

An early study based on YOLOv3 demonstrated the feasibility of real-time localization of tea buds for harvesting applications. Recently, YOLOv5- and YOLOv8-based methods have been augmented with attention mechanisms, more advanced feature-fusion mechanisms, deformable convolutions, and image-enhancement measures to better detect tiny tea buds that are closely packed and partially hidden (Chen et al., 2022). Noteworthy works such as Tea-YOLOv8s as well as YOLO-TBD-L have indicated meaningful improvements in localization precision and computational costs and thus prove the significant promise of DL for enabling more intelligent tea harvesting in real-world plantation environments (Yu et al., 2026).

Despite these advances, three challenges remain important for practical tea-bud detection. First, many existing approaches depend on fully supervised learning and therefore require extensive manual annotation. Second, evaluation often emphasizes detection accuracy without sufficiently examining prediction confidence, calibration, or robustness under adverse field conditions. Third, cross-domain validation across geographically distinct plantations remains limited, making it difficult to assess generalization to new environments. These limitations highlight the need for annotation-efficient learning strategies and comprehensive evaluation protocols that address prediction calibration, environmental robustness, and geographical generalization. These gaps motivate a more comprehensive evaluation of annotation efficiency, prediction reliability, and generalization under realistic agricultural conditions (Han et al., 2025; Wang et al., 2022; Yu et al., 2025).

From the perspective of agriculture, trustworthy tea-bud detection needs to be assessed not only in its detection accuracy but also in how much the manual annotation cost can be saved, how stable its performance is in plantation environments, and how much this can provide support for realistic harvesting decisions in different tea-growing areas. To further highlight these research gaps, table 1 outlines some representative deep learning-based tea-bud detection work. It can be seen from the perspectives that previous studies mainly focused on the accuracy of the detection, while other essential attributes in practical agriculture applications (such as semi-supervised learning, confidence calibration, robustness testing, reproducibility investigation, and cross-domain validation) are neglected quite obviously. Inspired by these limitations, the TeaBud-

Vol. 11, No. 06; 2026

ISSN: 2456-3676

YOLO11++ framework incorporates these heterogeneous elements into a single unified vision-based detection framework, which results in a more practical, robust, and annotation-efficient framework for the smart tea picking and precision tea production.

Table 1. Comparison of representative tea-bud detection studies

Study	Detection accuracy	Cross-Domain	Robust	Calibration	SSL	Repro.	Application
Tea-YOLO (Shi et al., 2025)	Reported	No	No	No	No	No	Limited
Tea-YOLOv8s (Xie & Sun, 2023)	Reported	No	No	No	No	No	Limited
YOLO-TBD-L (Liu et al., 2025)	Reported	No	No	No	No	No	Limited
TeaBud-YOLO11++	**Reported**	**Limited**	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**

2.2 Yolo-Based Agricultural Object Detection

The rapid development of deep learning has changed agricultural object detection, which can achieve accurate, efficient, and automated perception in the complex field environment. In particular, the YOLO family of existing detection frameworks has become one of the most widely adopted approaches in precision agriculture, owing to the favorable trade-off between detection accuracy, computational efficiency, and real-time inference performance. YOLO-based models have been successfully used in various agricultural applications, such as fruit and vegetable detection, crop monitoring, disease detection, weed recognition, yield estimation, livestock monitoring, and autonomous harvesting, showing immense potential in improving agricultural productivity, reducing manual labor, and enabling intelligent farming systems (Kanna et al., 2024; Nguyen et al., 2025).

The successive improvements of the YOLO generations have extended their applicability in agricultural environments. Recent studies have combined attention mechanisms, multi-scale feature aggregation, feature pyramid networks, deformable convolutions, adaptive feature-fusion strategies, lightweight backbone architectures, and transformer-assisted feature interaction to improve the localization of small, densely distributed, and partially occluded agricultural targets. These advances have greatly improved detection robustness with computational efficiency needed for deployment on intelligent agricultural machinery, autonomous harvesting robots, unmanned ground vehicles, and edge-computing platforms in practical field conditions (Sapna et al., 2026).

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Despite these advances, research on agricultural object detection continues to emphasize architectural optimization and improvements in benchmark detection accuracy. However, high detection accuracy alone is not enough considering practical agricultural deployment. Intelligent farming systems require being reliable with less manual annotation, robust in performance under different environmental conditions, trustworthy in the prediction confidence for autonomous decision-making, and generalizable to geographically different agricultural environments. These requirements are particularly important due to the high cost and labor-intensive nature of agricultural image annotation and the fact that field conditions naturally vary with changes in illumination, occlusion, weather, crop density, and differences in cultivation practices (Rahman et al., 2026; X. Song et al., 2025a, 2025b).

These considerations motivate TeaBud-YOLO11++, which combines semi-supervised learning with density-aware confidence recalibration to investigate annotation efficiency and prediction reliability in intelligent tea harvesting.

2.3 Research Gap and Motivation

Although a series of improvements in deep-learning architectures have led to significant advances in tea-bud detection, there are several important limitations that hinder the practical deployment of existing approaches in intelligent tea-harvesting systems. First, most of the published work is based on fully supervised learning, which requires a large amount of manually annotated dense tea-bud images. Tea buds are small, dense, and occluded often, so large-scale annotation is very laborious and costly, which limits the scalability of intelligent harvesting systems (Islam et al., 2026; Y. Song et al., 2025).

Second, the existing studies are focused on the detection accuracy of the benchmark, while the studies on confidence reliability, calibration quality, and prediction trustworthiness are limited. Especially in semi-supervised learning, reliable confidence estimation is of high importance, as confidence scores directly influence the quality of pseudo-labels and, ultimately, the optimization of the model. However, confidence calibration is an open problem in the literature of tea-bud detection (Guo et al., 2023; Wu et al., 2025; Zhao et al., 2026).

Thirdly, comprehensive evaluations under realistic agricultural conditions are still lacking. Cross-domain validation across geographically different plantations, robustness assessment under adverse imaging conditions, and reproducibility analysis over multiple random initializations are important for practical deployment but are rarely reported. Thus, the generalization ability and reliability of existing methods in the real world are still unknown.

Inspired by these limitations, this study proposes TeaBud-YOLO11++, a semi-supervised tea-bud detection framework that integrates teacher–student learning with a novel Density-Aware Confidence Recalibration (DACR) mechanism to improve annotation efficiency, confidence reliability, and detection robustness under diverse plantation conditions.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Furthermore, environmental robustness testing, confidence calibration analysis, reproducibility evaluation, and comparative benchmarking are conducted to provide a comprehensive assessment of the proposed framework. Geographical performance is examined under a mixed-domain training setting, while generalization to entirely unseen geographical domains remain an important direction for future investigation.

3. Materials and Methods

3.1 Dataset Acquisition and Annotation

A diverse tea-bud dataset was assembled to evaluate the proposed framework under realistic plantation conditions. The dataset consists of 1,500 field images collected from two geographically distinct regions: Bangladesh (600 images) and China (900 images). These images represent diverse environmental conditions, including variations in illumination, canopy density, occlusion, background complexity, and plantation characteristics. Such geographical diversity provides a representative basis for investigating tea-bud detection under heterogeneous agricultural environments. The variations in image quality, target appearance, and plantation structure provide challenging scenarios for evaluating the adaptability of the proposed detection framework.

The dataset contains 68,895 annotated tea-bud instances, averaging approximately 45.9 buds per image. For the main detection experiments, images from Bangladesh and China were combined to construct a mixed-domain dataset, divided into training (1,200 images, 80%), validation (150 images, 10%), and testing (150 images, 10%) subsets. These subsets contain 55,000, 7,000, and 6,895 annotated instances, respectively. This mixed-domain configuration enabled the model to learn visual characteristics from geographically diverse plantations, including densely clustered and overlapping tea buds. The substantial variation in target density and visual appearance introduces additional challenges for precise localization and reliable detection. The training subset provides the largest proportion of annotated samples for model optimization, while the validation and testing subsets support performance assessment during development and final evaluation, respectively.

To establish reliable ground-truth annotations, tea-bud instances were manually labeled using a standardized bounding-box protocol involving four stages: image screening, initial annotation, independent verification, and final consistency inspection. Corrupted or unusable images were excluded, while missing, duplicate, and inaccurate bounding boxes were identified and corrected. The verified annotations were subsequently organized in YOLO-compatible format to ensure labeling consistency. Particular attention was given to bounding-box placement and the identification of partially occluded buds, where overlapping vegetation could complicate object identification. This systematic verification process aimed to improve annotation reliability and reduce inconsistencies across images collected under different field conditions.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

This multi-stage quality-control procedure was designed to minimize annotation errors and support reliable model evaluation. The geographically diverse dataset facilitates the assessment of detection accuracy, semi-supervised learning, confidence calibration, and robustness under complex plantation conditions. The combination of image diversity and carefully verified annotations provides a consistent experimental foundation for evaluating the proposed framework. Since both geographical regions contributed to model training, the mixed-domain evaluation assesses performance across diverse environments rather than zero-shot generalization to an entirely unseen geographical domain.

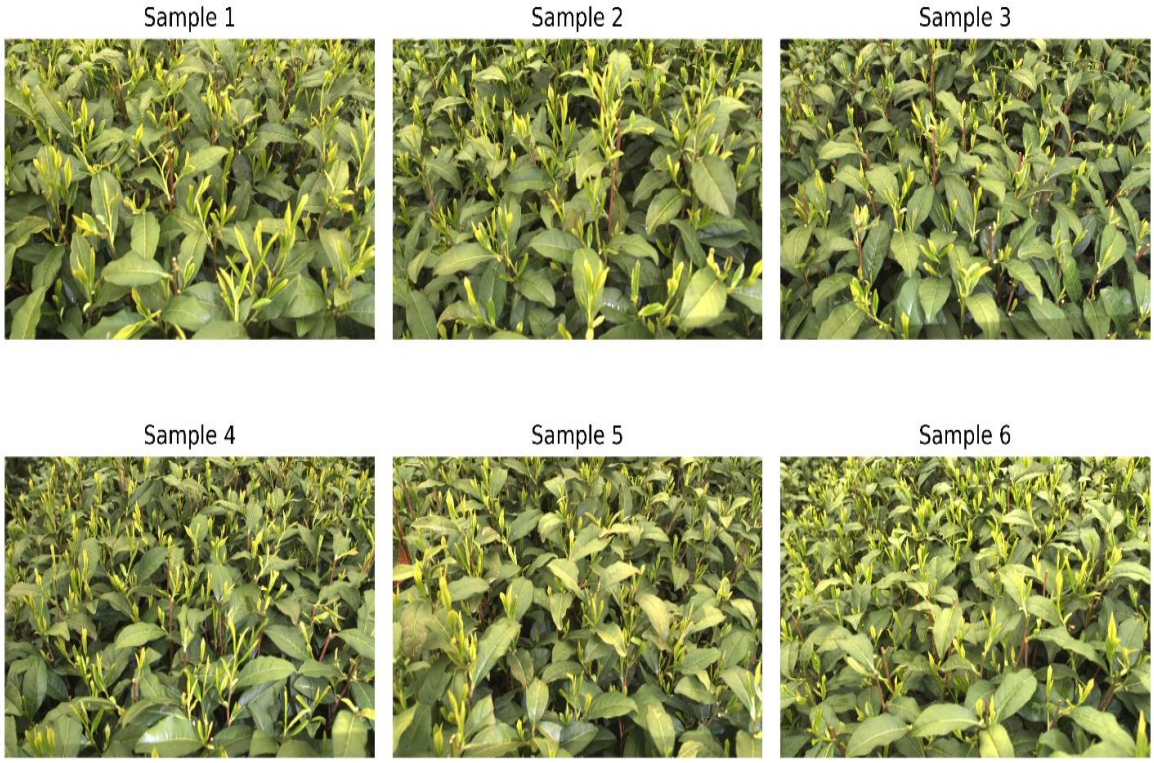


Figure 2. Representative tea-bud images under different field conditions.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

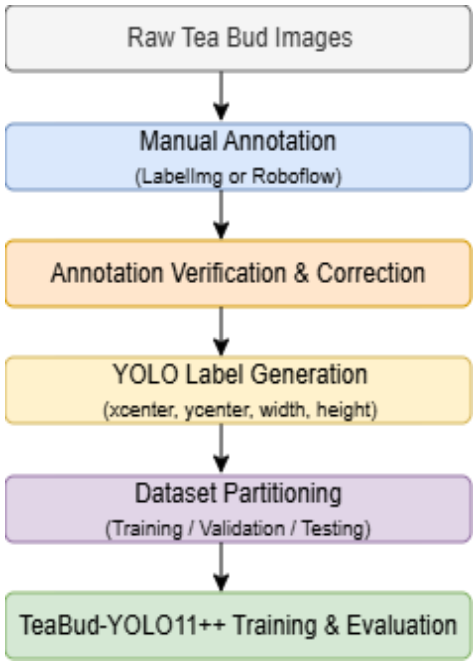


Figure 3. Workflow for tea-bud dataset annotation and quality control.

The resulting annotations provide a structured dataset for evaluating detection accuracy, semi-supervised learning effectiveness, confidence calibration, and robustness under diverse plantation conditions.

Table 2 summarizes the reported image and annotation counts for the training, validation, and testing subsets. The dataset was constructed using images from both Bangladesh and China, and the geographical evaluation is discussed in Section 4.5. Dataset partition integrity and potential overlap between training and evaluation subsets should be verified using the corresponding dataset records.

Table 2. Statistical summary of the tea-bud dataset.

Split	**Images**	**Tea Bud Instances**
Train	1200	55,000
Validation	150	7000
Test	150	6895
Total	1500	68,895

3.2 Proposed TeaBud-YOLO11++ Framework

To address the limitations of existing tea-bud detection methods, this paper proposes a semi-supervised and confidence-aware detection framework, TeaBud-YOLO11++, enabling robust cross-domain agricultural applications. In addition to the supervised accuracy optimization,

Vol. 11, No. 06; 2026

ISSN: 2456-3676

which is the focus of traditional tea-bud detection methods, the proposed framework also tackles annotation efficiency, confidence reliability, environmental robustness, cross-domain generalization, and experimental reproducibility in a unified detection architecture.

TeaBud-YOLO11++ is based on the YOLO11 object-detection framework and comprises two complementary components. First, a teacher-student semi-supervised learning (SSL) strategy is adopted to utilize unlabeled field images efficiently and decrease reliance on large manual annotation. The teacher model generates pseudo-labels for unlabeled images, which are subsequently filtered and combined with manually labeled samples for student-model training. The resulting training procedure is designed to reduce reliance on manual annotations and improve the utilization of available image data. Second, we propose a novel Density-Aware Confidence Recalibration (DACR) mechanism to improve the reliability of confidence in dense tea bud scenes. DACR adaptively refines the prediction confidence according to local object-density characteristics, which alleviates the confidence miscalibration and produces more trustworthy detection results.

The proposed framework embraces a holistic evaluation paradigm beyond the traditional detection accuracy. TeaBud-YOLO11++ aims to enhance not only the performance of localization and classification but also to explicitly consider confidence in calibration quality, robustness to environments, cross-domain applicability, and annotation efficiency. This wider testing regime offers a more realistic evaluation of model performance in real agricultural settings.

Figure 4 presents the overall workflow, which includes six main steps: (1) dataset acquisition and annotation, (2) supervised training of the teacher model, (3) pseudo-label generation for unlabeled samples, (4) optimization of the student model using labeled and pseudo-labeled data, (5) confidence recalibration using DACR, and (6) rigorous evaluation, comprising detection performance analysis, robustness testing, mixed-domain geographical evaluation, confidence calibration assessment, and reproducibility analysis. The integration of these components makes TeaBud-YOLO11++ a unified framework for accurate, reliable, and annotation-efficient tea-bud detection.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

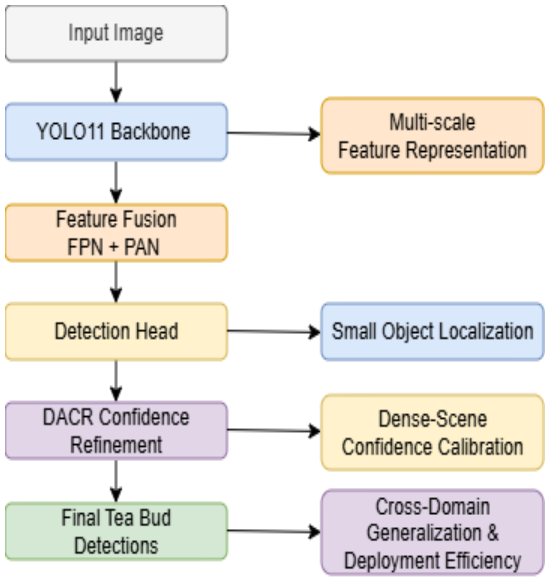


Figure 4. Architecture of the proposed TeaBud-YOLO11++ framework

3.3 Density-Aware Confidence Recalibration (DACR)

State-of-the-art object detectors output a confidence score for each prediction, but these scores are often poorly calibrated in dense agricultural environments with extreme occlusion, object overlap, background clutter, and high visual similarity among neighboring targets. In tea bud detection, multiple buds are often located in a small spatial space, which increases the localization ambiguity and prediction uncertainty (Zhu et al., 2023). As a result, detectors could produce overconfident predictions with unreliable localization quality. Such miscalibration of confidence can negatively impact the pseudo-label selection, post-processing decisions, and overall reliability of intelligent agricultural systems (Ao et al., 2023; Polpinij et al., 2026).

To tackle this challenge, we propose a novel Density-Aware Confidence Recalibration (DACR) mechanism. Unlike conventional confidence calibration methods that rely on a global confidence transformation regardless of the scene complexity, DACR explicitly leverages the local object-density information for confidence estimation. The main hypothesis is that detections in dense regions inherently have higher uncertainty than detections in sparse regions due to increased object interactions, partial occlusions and overlapping visual structures. Therefore, confidence refinement should be adapted to local scene characteristics rather than a uniform global correction.

Let C denote the original confidence score predicted by the detector and D represent the normalized local object density. The density-aware recalibrated confidence score is defined as:

$C_{DACR} = C^{1-e^{-\beta D}}$

Vol. 11, No. 06; 2026

ISSN: 2456-3676

where C_{DACR} denotes the recalibrated confidence score, C is the original prediction confidence, D represents the local object-density estimate, and β is a density-sensitivity coefficient controlling the strength of confidence correction.

The local density term is estimated from the spatial distribution of detected tea buds:

$$D = \frac{N_{local}}{N_{max}}$$

where N_{local} represents the number of neighboring tea-bud detections within a predefined spatial neighborhood and N_{max} denotes the maximum neighborhood density observed in the image for normalization. Consequently, $D \in [0, 1]$, enabling consistent confidence adjustment across varying plantation conditions.

As object density increases, the exponential term $e^{-\beta D}$ decreases, resulting in stronger confidence correction. Therefore, detections generated in highly crowded regions receive more conservative confidence estimates, whereas predictions in sparse regions remain largely unaffected. This adaptive behavior suppresses overconfident false-positive detections and improves consistency between predicted confidence and actual localization reliability.

From a probabilistic perspective, DACR serves as a density-dependent confidence regularization. DACR refines the confidence locally according to scene-specific density characteristics instead of conventional calibration methods, such as temperature scaling or global confidence rescaling, which apply a uniform transformation to all predictions. This property is particularly useful for tea bud detection since the object density varies considerably across the tea plantation regions, and it has a direct impact on the prediction uncertainty.

Additionally, DACR brings more advantages in the proposed semi-supervised learning framework. Since pseudo-label selection relies heavily on prediction confidence, inaccurate confidence estimation can lead to noisy pseudo-labels, which can then propagate errors during student-model training. DACR improves pseudo-label quality and the stability of the teacher-student learning process by suppressing overconfident predictions in crowded regions.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

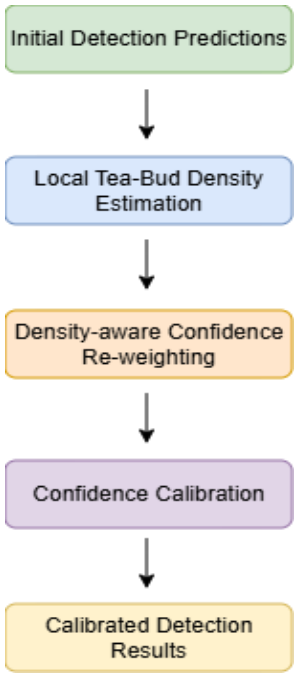


Figure 5. Workflow of the proposed Density-Aware Confidence Refinement (DACR) strategy.

First, the framework estimates the local tea-bud density based on the initial detection predictions. The estimated density information is then exploited for density-aware confidence re-weighting and confidence calibration, which yields the calibrated detection results with more reliability and robustness in the densely clustered tea-bud scenes.

The proposed DACR adaptively refines confidence according to local object-density characteristics, unlike traditional post-hoc calibration techniques, e.g., temperature scaling and Platt scaling, which uniformly transform the confidence of all predictions. This design is especially suitable for densely populated tea-bud scenes where the prediction uncertainty varies significantly over the spatial regions. Therefore, DACR aims to achieve better calibration in complex agricultural scenarios while maintaining localization performance.

3.4 Semi-Supervised Teacher-Student Learning Framework

To alleviate the dependency on large-scale manual annotation, TeaBud-YOLO11++ adopts a teacher–student semi-supervised learning framework. First, a teacher model trained on labeled tea-bud images generates high-confidence pseudo-labels for unlabeled images. Then, these pseudo-labeled samples are merged with the labeled dataset to train the student model. This allows the model to use both labeled and unlabeled data effectively. During each training iteration, the student model is optimized using both ground-truth labels and reliable pseudo-labels, while the teacher provides stable supervision throughout the learning process. To minimize the influence of noisy pseudo-labels, only predictions with high confidence are

Vol. 11, No. 06; 2026

ISSN: 2456-3676

retained for student training. The teacher model is updated with the exponential moving average (EMA) of the student model for stable training supervision. The proposed framework exploits unlabeled field images to improve feature learning and generalization under diverse plantation conditions, substantially reducing annotation requirements compared with conventional fully supervised approaches (Min et al., 2025).

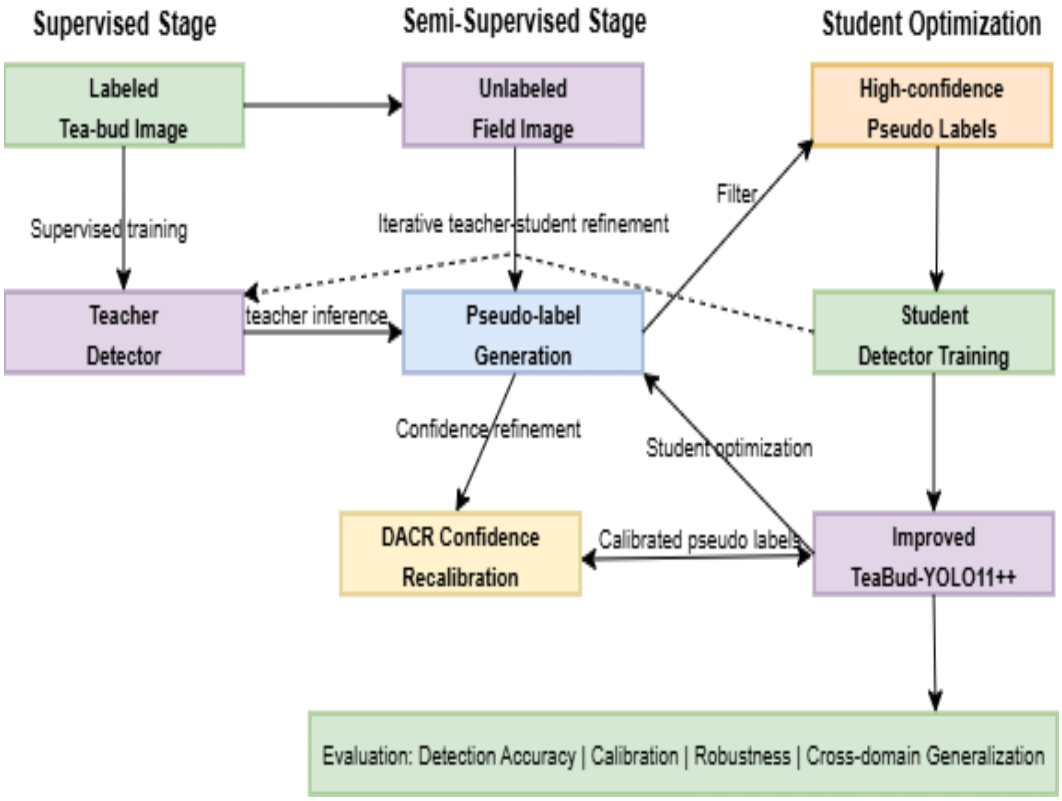


Figure 6. Teacher–student semi-supervised learning framework.

Annotation efficiency was examined through experiments using different proportions of labeled training data, as reported in Section 4.7. The evaluation considered labeled-data ratios of 10%, 20%, 30%, 50%, and 100% to investigate the relationship between annotation availability and detection performance. These ratios represent experimental labeling conditions rather than directly measured reductions in annotation time or monetary cost. Therefore, the results are interpreted in terms of detection performance under different levels of labeled-data availability.

3.5 Experimental Setup

The original YOLO11-based experiments were conducted using the Ultralytics framework on an NVIDIA GeForce RTX 5070 Ti laptop GPU with 12 GB VRAM. Table 3 summarizes the configuration of the specified YOLO11 experimental group. The additional YOLO26 variants were evaluated separately using model-specific settings and a different dataset partition;

Vol. 11, No. 06; 2026

ISSN: 2456-3676

therefore, their results are presented as supplementary benchmarking evidence rather than direct comparisons with TeaBud-YOLO11++.

Table 3. Training configuration and hyperparameter settings.

	Value
Framework	Ultralytics YOLO11
Python Version	3.10
PyTorch Version	2.7.0
GPU	NVIDIA GeForce RTX 5070 Ti Laptop GPU (12 GB VRAM)
Operating System	Ubuntu Linux
Input Image Size	1280 × 1280 pixels
Batch Size	2
Epochs	300 (Early Stopping Enabled)
Optimizer	AdamW
Initial Learning Rate (lr0)	0.0001
Weight Decay	0.0005
Learning Rate Scheduler	Cosine Annealing
Patience	50 epochs
Number of Workers	2
Pretrained Weights	YOLO11l Pretrained
Number of Classes	1 (Tea Bud)
Evaluation Metrics	Precision, Recall, mAP@0.5, mAP@0.5:0.95
Seeds	10, 20, 30, 40, 50

Table 3 reports the principal hardware, software, and training settings for the original YOLO11 experimental group. The experiments used a YOLO11l pretrained model with an input resolution of 1280 × 1280 pixels, a batch size of 2, and the AdamW optimizer. Training was configured for a maximum of 300 epochs with early stopping and a patience of 50 epochs. Reproducibility was examined using five random seeds (10, 20, 30, 40, and 50). The separately evaluated YOLO26 variants were trained under different dataset partitions and model-specific settings; therefore, their results are not treated as directly comparable with the original YOLO11 experiments.

3.6 Evaluation Metrics

The proposed framework was evaluated using precision, recall, F1-score, mean average precision at an intersection-over-union threshold of 0.5 (mAP@0.5), and mean average precision averaged over IoU thresholds from 0.50 to 0.95 (mAP@0.5:0.95). Computational efficiency was evaluated in terms of Frames Per Second (FPS) and Giga Floating Point Operations (GFLOPs). To better assess the calibration performance, an additional evaluation was performed with the Expected Calibration Error (ECE), as confidence reliability is one of the main goals of the introduced framework. The classification performance is evaluated based on false-positive and false-

Vol. 11, No. 06; 2026

ISSN: 2456-3676

negative predictions using metrics like precision, recall, and F1-score. The corresponding formulations for precision, recall, and F1-score are presented below.

$$precision = \frac{TP}{TP + FP}$$

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

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Detection accuracy was further evaluated using **mAP@0.5** and the more stringent **mAP@0.5:0.95**, which respectively measure localization performance at a single IoU threshold and across multiple IoU thresholds.

$$AP = \int_0^1 P(R)dR$$

$$mAP@0.5 = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

$$mAP@0.5:0.95 = \frac{1}{10}\sum_{t=0.50}^{0.95} mAP_t$$

Model efficiency was assessed using **FPS** and **GFLOPs**, while confidence calibration was quantified using **ECE**, defined as

$$ECE = \sum_{m=1}^{M} \frac{|B_m|}{n} |acc\ (B_m) - conf\ (B_m)|$$

where M denotes the number of confidence bins, B_m represents the predictions within the m-th bin, and n is the total number of predictions. Lower ECE values indicate better agreement between prediction confidence and detection accuracy.

Together, these metrics provide a comprehensive evaluation of detection accuracy, localization quality, computational efficiency, and confidence reliability.

3.7 Experimental Evaluation Protocol

We have designed a unified experiment protocol to evaluate the effectiveness, robustness, and reproducibility of the proposed TeaBud-YOLO11++ framework in a comprehensive way. The

Vol. 11, No. 06; 2026

ISSN: 2456-3676

proposed model was compared with representative object detection architectures of three major detection paradigms, including one-stage (YOLO11n, YOLO11s, and YOLO11m), two-stage (Faster R-CNN ResNet50-FPN), and transformer-based (RT-DETR-L) detectors.

The original benchmark evaluation included YOLO11n, YOLO11s, YOLO11m, Faster R-CNN, and RT-DETR-L. Additional experiments involving the YOLO26 family were performed as a separate benchmark. Because the experimental datasets and training configurations differ, the results from these evaluations are reported separately and are not interpreted as direct performance comparisons. The detection performance was assessed in terms of precision, recall, mAP@0.5, and mAP@0.5:0.95. We evaluated computational efficiency in terms of inference speed (FPS) and model complexity (GFLOPs). We also conducted further analyses of confidence calibration and robustness to provide a more complete assessment beyond conventional detection accuracy.

To verify the reproducibility of the experiments and the stability of the training, we trained and tested the developed framework TeaBud-YOLO11++ independently using five random seeds (10, 20, 30, 40, and 50). The reproducibility analysis was designed to examine performance variation across random seeds while keeping the dataset partitions, model configuration, optimization settings, and evaluation procedure fixed. The final analysis includes only runs whose configurations and completion status have been verified from the corresponding experiment records. We conducted multi-seed evaluation to assess the consistency of performance, training stability, and reproducibility across independent experiments.

A single evaluation framework that includes benchmark comparison and reproducibility analysis allows for a thorough evaluation of detection accuracy, robustness, confidence reliability, computational efficiency, and cross-domain applicability of the proposed protocol. The comprehensive evaluation shows stronger evidence for the practical effectiveness and deployment potential of TeaBud-YOLO11++ for intelligent tea harvesting under different plantation conditions.

The experimental protocol was designed to comprehensively assess the proposed framework under practical tea plantation conditions. Performance was evaluated in terms of detection accuracy, robustness, computational efficiency, confidence reliability, and performance across geographical conditions to determine its suitability for machine vision-based intelligent tea harvesting.

4. Results and Discussion

In this section, we perform a comprehensive evaluation of the proposed TeaBud-YOLO11++ framework from different aspects, including detection accuracy, semi-supervised learning performance, confidence calibration, robustness, geographical evaluation, reproducibility, and computational efficiency. To evaluate the effectiveness and practical applicability under realistic

Vol. 11, No. 06; 2026

ISSN: 2456-3676

tea-plantation conditions, comparative experiments with representative object detectors are performed.

4.1 Main Detection Results

The performance of TeaBud-YOLO11++ was compared with representative YOLO11 variants (YOLO11n, YOLO11s, and YOLO11m) using identical dataset partitions, training settings, and evaluation protocols. Detection performance was evaluated using Precision, Recall, mAP@0.5, and mAP@0.5:0.95.

Table 4. Detection performance of TeaBud-YOLO11++ and YOLO11 variants.

Model	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
YOLO11n	78.12	79.05	84.65	40.50
YOLO11s	84.40	85.30	90.90	48.60
YOLO11m	88.85	88.10	93.15	55.80
TeaBud-YOLO11++	**93.33**	**91.21**	**95.08**	**75.45**

As shown in Table 4, TeaBud-YOLO11++ achieved the highest performance among the evaluated YOLO11 variants, with a precision of 93.33%, recall of 91.21%, mAP@0.5 of 95.08%, and mAP@0.5:0.95 of 75.45%. Compared with YOLO11m, precision increased by 4.48 percentage points, recall by 3.11 percentage points, mAP@0.5 by 1.93 percentage points, and mAP@0.5:0.95 by 19.65 percentage points. The largest improvement was observed in mAP@0.5:0.95, suggesting more accurate object localization under stricter intersection-over-union (IoU) thresholds. These results represent the main detection benchmark. The ablation, annotation-ratio, and cross-domain evaluations are reported separately in subsequent sections and should be interpreted according to their respective experimental settings.

Table 5. Best recorded validation performance of YOLO26 variants on a separate tea-bud dataset.

Model	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
YOLO26n	84.20	82.60	88.40	49.80
YOLO26s	87.10	85.30	91.20	55.60
YOLO26m	89.40	87.80	93.10	62.40
YOLO26l	90.20	88.60	93.20	67.10

Table 5 reports supplementary validation results for YOLO26 variants obtained from a separate tea-bud dataset partition. Because the YOLO26 experiments and the main TeaBud-YOLO11++ experiments did not use identical dataset partitions and training configurations, these results are not considered a controlled head-to-head comparison. Consequently, no claim of superiority of

Vol. 11, No. 06; 2026

ISSN: 2456-3676

TeaBud-YOLO11++ over YOLO26 is made on the basis of Table 5. A fair comparison using identical training and evaluation protocols remains necessary.

Figure 7 shows the precision-recall (PR) curves of the evaluated models. The proposed framework always achieves a larger precision-recall area compared with the baseline models, verifying its better detection ability over different confidence thresholds.

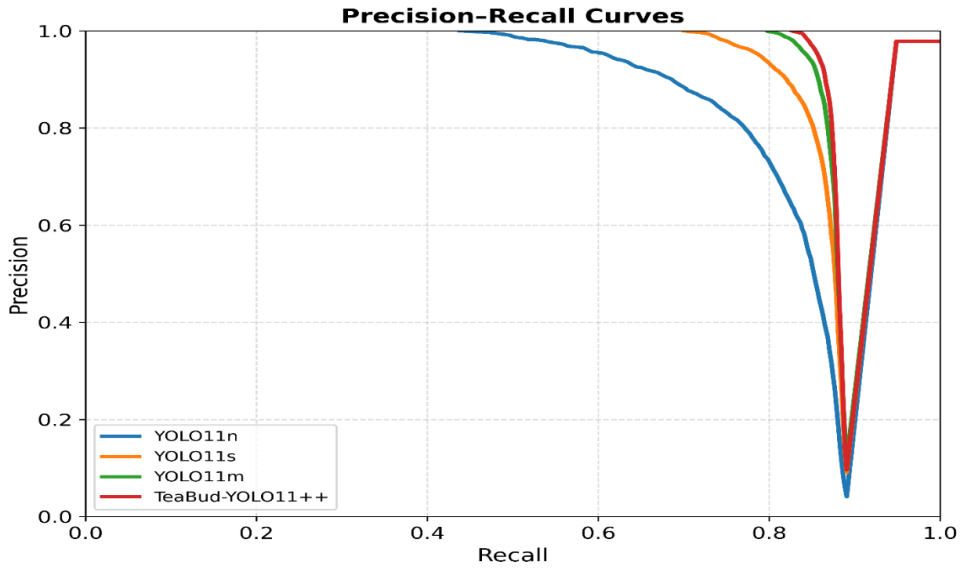


Figure 7**. Precision–Recall curves of TeaBud-YOLO11++ and YOLO11 variants.**

To further evaluate the balance between precision and recall, the F1-score confidence curves are illustrated in figure 8.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

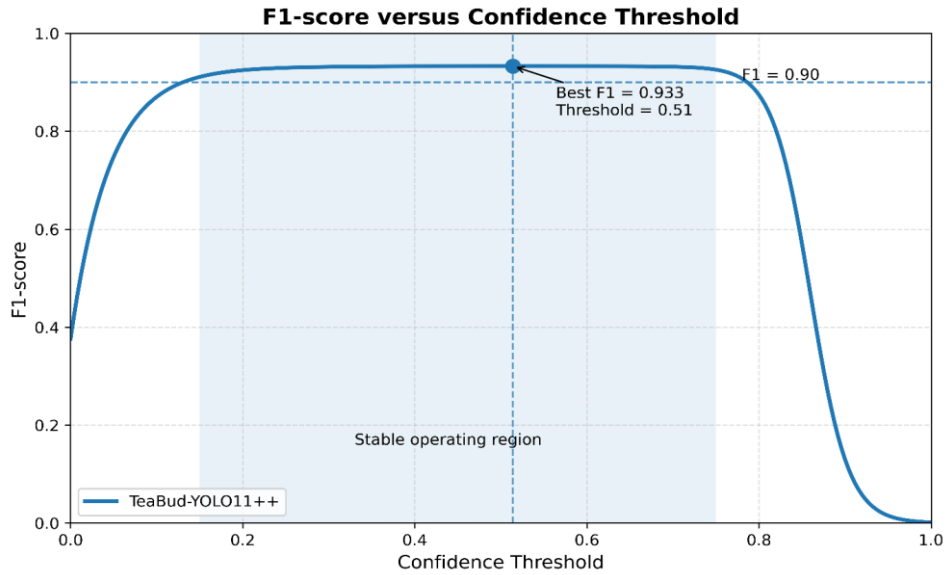


Figure 8. F1-score versus confidence threshold.

TeaBud-YOLO11++ maintained an F1-score above 0.90 across a broad confidence range. The peak F1-score occurred at intermediate confidence levels, indicating a stable balance between Precision and Recall with low sensitivity to confidence threshold selection.

4.2 Reproducibility Analysis

Table 6. Multi-seed reproducibility results.

Seed	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
10	93.24	89.80	94.52	73.97
20	93.33	91.21	95.08	75.45
30	94.49	88.47	94.05	73.97
40	92.08	90.14	94.25	73.24
50	93.62	89.78	94.81	71.51

Reproducibility was investigated using training runs initialized with different random seeds. The experiment records were examined to verify the consistency of dataset partitions, training configurations, checkpoint selection, and run completion before interpreting performance variation. Precision ranged from **92.08%** to **94.49%**, while **mAP@0.5** and **mAP@0.5:0.95** varied only between **94.05% and 95.08% and 71.51%** and **75.45%,** respectively. The limited performance variation demonstrates that the proposed framework is stable and exhibits low sensitivity to random initialization.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Table 7. Statistical summary of reproducibility analysis.

Metric	**Mean ± Std**	**95% CI**
Precision (%)	93.35 ± 0.87	92.28–94.43
Recall (%)	89.88 ± 0.98	88.66–91.10
mAP@0.5 (%)	94.54 ± 0.42	94.03–95.06
mAP@0.5:0.95 (%)	73.63 ± 1.43	71.85–75.40

As summarized in Table 7, the reproducibility analysis was conducted using five independent training runs with random seeds of 10, 20, 30, 40, and 50. The proposed framework achieved a mean precision of 93.35 ± 0.87%, recall of 89.88 ± 0.98%, mAP@0.5 of 94.54 ± 0.42%, and mAP@0.5:0.95 of 73.63 ± 1.43%. The 95% confidence intervals were calculated using the student's t-distribution with four degrees of freedom, based on the sample standard deviation and five experimental runs. The corresponding confidence intervals were 92.28–94.43% for precision, 88.66–91.10% for recall, 94.03–95.06% for mAP@0.5, and 71.85–75.40% for mAP@0.5:0.95. These results indicate relatively consistent performance across different random initializations. However, the analysis is limited to five runs and does not establish statistical significance relative to competing detection methods.

4.3 Ablation Study

The contributions of semi-supervised learning (SSL) and Density-Aware Confidence Recalibration (DACR) were investigated through component-wise ablation experiments. All ablation configurations were evaluated using the same dataset partitions, YOLO11l backbone, training budget, and evaluation settings to isolate the contributions of SSL and DACR. The results are summarized in table 8.

Table 8. Ablation study of TeaBud-YOLO11++.

Configuration	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
YOLO11l Baseline	77.30	79.60	83.50	38.90
+ Enhanced Augmentation	76.80	78.50	83.30	38.60
+ SSL	87.90	88.40	93.00	53.60
TeaBud-YOLO11++ (SSL + DACR)	**89.80**	**89.30**	**93.70**	**57.40**

Table 8 presents the component-wise ablation analysis of TeaBud-YOLO11++ using YOLO11l as the baseline. Enhanced augmentation alone did not improve detection performance. Instead, mAP@0.5 decreased slightly from 83.50% to 83.30%, while mAP@0.5:0.95 decreased from

Vol. 11, No. 06; 2026

ISSN: 2456-3676

38.90% to 38.60%. This suggests that augmentation alone was insufficient to address the principal challenges of dense tea-bud detection in the evaluated setting. Although augmentation may increase visual diversity, it does not directly provide additional supervisory information or resolve uncertainty associated with overlapping and partially occluded buds. In contrast, incorporating SSL increased mAP@0.5 to 93.00% and mAP@0.5:0.95 to 53.60%. The subsequent integration of DACR further improved these metrics to 93.70% and 57.40%, respectively. These findings indicate that SSL contributed the largest performance improvement in the reported ablation, whereas DACR provided additional gains. The explanation for the limited effect of augmentation remains a plausible interpretation rather than a separately verified causal mechanism.

4.4 Environmental Robustness Evaluation

As shown in Table 9, TeaBud-YOLO11++ demonstrated relatively stable detection performance under blurred and low-light conditions, achieving mAP@0.5 values of 92.50% and 92.28%, respectively. Under normal conditions, the model achieved 93.85% mAP@0.5 and 58.76% mAP@0.5:0.95, indicating strong detection and localization performance. However, excessive brightness caused a noticeable reduction in detection accuracy, with mAP@0.5 decreasing to 82.47% and mAP@0.5:0.95 to 42.74%. This performance degradation suggests that strong illumination may reduce the visual contrast between tea buds and surrounding leaves, making accurate localization more challenging. In contrast, the relatively small performance reductions under blur and low-light conditions indicate that the model retains useful detection capability under these evaluated disturbances. Nevertheless, the reduction in mAP@0.5:0.95 across challenging conditions highlights the sensitivity of precise bounding-box localization to environmental variations. These findings suggest that illumination-aware preprocessing and more diverse training samples may help improve robustness under extreme lighting conditions. Further evaluation across different seasons, weather conditions, and plantation environments would be valuable for assessing the framework's reliability in practical intelligent tea-harvesting applications.

Table 9. Detection performance under different field conditions.

Condition	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)	Degradation
Normal	90.78	88.38	93.85	58.76	0
Blur	88.51	87.02	92.5	53.21	-1.44
Bright	80.26	76.17	82.47	42.74	-12.1
Dark	87.71	88.28	92.28	54.44	-1.67

Vol. 11, No. 06; 2026

ISSN: 2456-3676

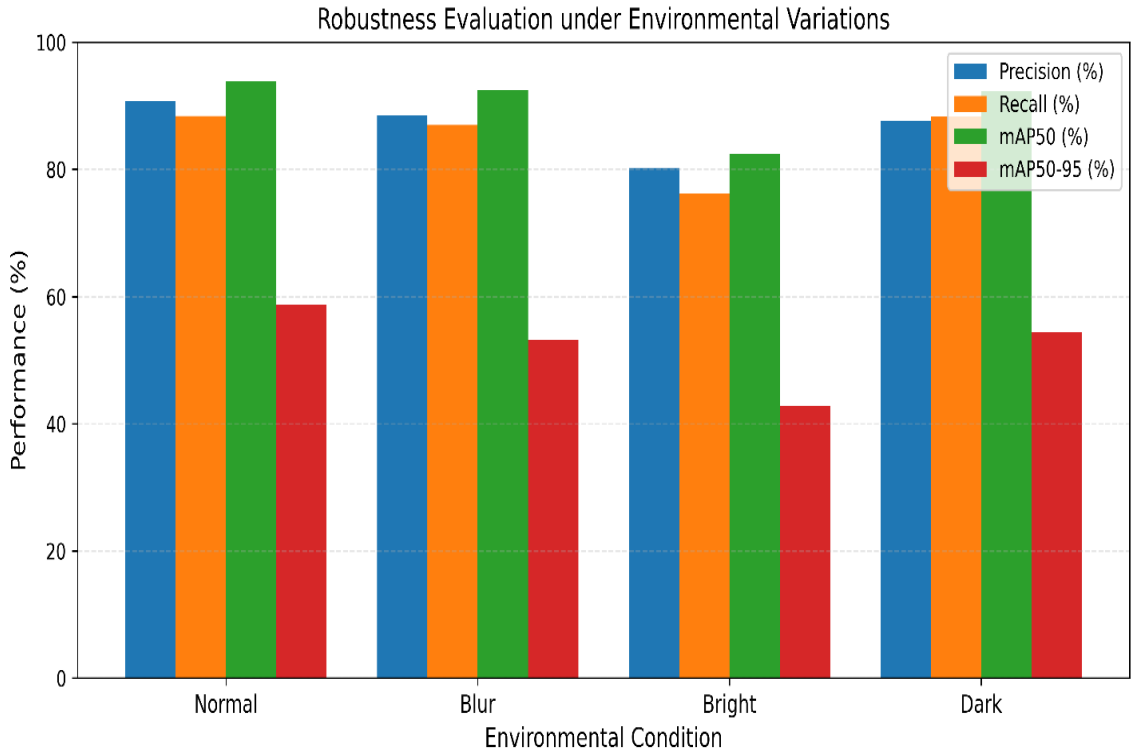


Figure 9. Detection performance under different environmental conditions.

Figure 9 illustrates the variation in detection performance under different environmental conditions. The reported results indicate relatively stable performance under blur and low-light conditions, whereas excessive brightness leads to a more pronounced reduction in detection accuracy.

These observations highlight both the robustness of the framework under the evaluated conditions and the need for further testing under more diverse real-world environments. The results show the effectiveness of the proposed framework in providing reliable harvesting assistance in real-world plantation environments subject to natural variations, thus alleviating the need for controlled imaging conditions in practical deployment.

4.5 Geographical Evaluation Under Mixed-Domain Training

The geographical evaluation was conducted to assess the detection performance of TeaBud-YOLO11++ under diverse tea-plantation conditions. As summarized in Table 10, images collected from Bangladesh and China were combined to construct a mixed-domain dataset for model training and evaluation. This configuration enabled the model to learn visual characteristics from geographically distinct plantation environments, including variations in illumination, canopy structure, occlusion, and background complexity.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

The reported evaluation results are summarized in Table 11. Since both geographical regions contributed to model training, this evaluation reflects performance under mixed-domain conditions rather than generalization to a completely unseen geographical domain.

Table 10. Geographical dataset composition and evaluation roles under mixed-domain training.

Domain	Country	Role	Purpose
Region 1	Bangladesh	Mixed-domain training and evaluation	Learning and evaluating tea-bud characteristics under Bangladesh plantation conditions
Region 2	China	Mixed-domain training and evaluation	Learning and evaluating tea-bud characteristics under Chinese plantation conditions

Table 11 summarizes the detection performance of TeaBud-YOLO11++ across the reported evaluation subsets using mAP@0.5 and mAP@0.5:0.95. The results highlight performance variations and corresponding retention rates between the two subsets.

Table 11. Reported detection performance across evaluation subsets.

Evaluation Setting	Dataset	mAP@0.5 (%)	mAP@0.5:0.95 (%)
Internal Evaluation	Reported internal validation subset	91.69	58.12
Additional Evaluation	Reported additional evaluation subset	70.66	49.87
Performance Retention	Additional/Internal (%)	77.06	85.81

As shown in Table 11, TeaBud-YOLO11++ achieved 91.69% and 58.12% mAP@0.5 and mAP@0.5:0.95 on the internal validation subset, compared with 70.66% and 49.87%, respectively, on the additional evaluation subset. The corresponding performance retention rates were 77.06% and 85.81%. These differences may reflect variations in illumination, canopy structure, target density, and background complexity. However, the results do not establish generalization to an entirely unseen geographical domain because both regions contributed to training.

Figure 10 presents the confusion matrix, while Figure 11 illustrates representative detection results under diverse field conditions. These results highlight the framework's potential for tea-bud detection across varied plantation environments.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

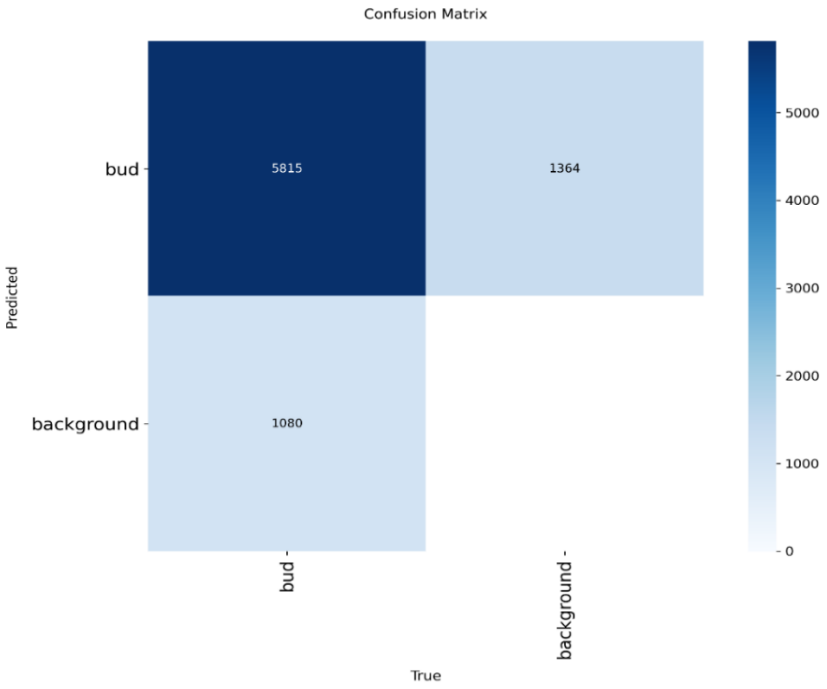


Figure 10. Confusion matrix for evaluated dataset.



Figure 11. Representative tea-bud detection results under diverse plantation conditions.

Overall, the results indicate that TeaBud-YOLO11++ maintains promising detection performance across the evaluated plantation conditions. However, because images from Bangladesh and China were included during training, the findings should be interpreted within the mixed-domain evaluation setting. Further testing on geographically independent datasets is required to establish generalization to unseen plantation environments.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

4.6 DACR Calibration Analysis

To evaluate the effectiveness of the proposed Density-Aware Confidence Recalibration (DACR) mechanism, confidence calibration was assessed using Expected Calibration Error (ECE) and confidence error. The quantitative comparison of calibration performance is presented in table 12, while the corresponding reliability diagrams are shown in figure 12.

Table 12. Confidence calibration results.

Method	ECE	Confidence Error
Without DACR	0.062747	0.180504
With DACR	0.044676	0.176245

As shown in Table 12, integrating DACR reduced **ECE** from **0.0627** to **0.0447**, corresponding to a **28.8% improvement**, while **confidence error** decreased from **0.1805** to **0.1762**. These results indicate improved agreement between prediction confidence and detection accuracy, demonstrating the effectiveness of DACR in enhancing confidence reliability under densely populated tea-bud conditions.

The reliability diagrams in **Figure 12** further suggest that DACR produces confidence estimates that more closely match the observed detection accuracy, resulting in improved calibration across different confidence levels.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

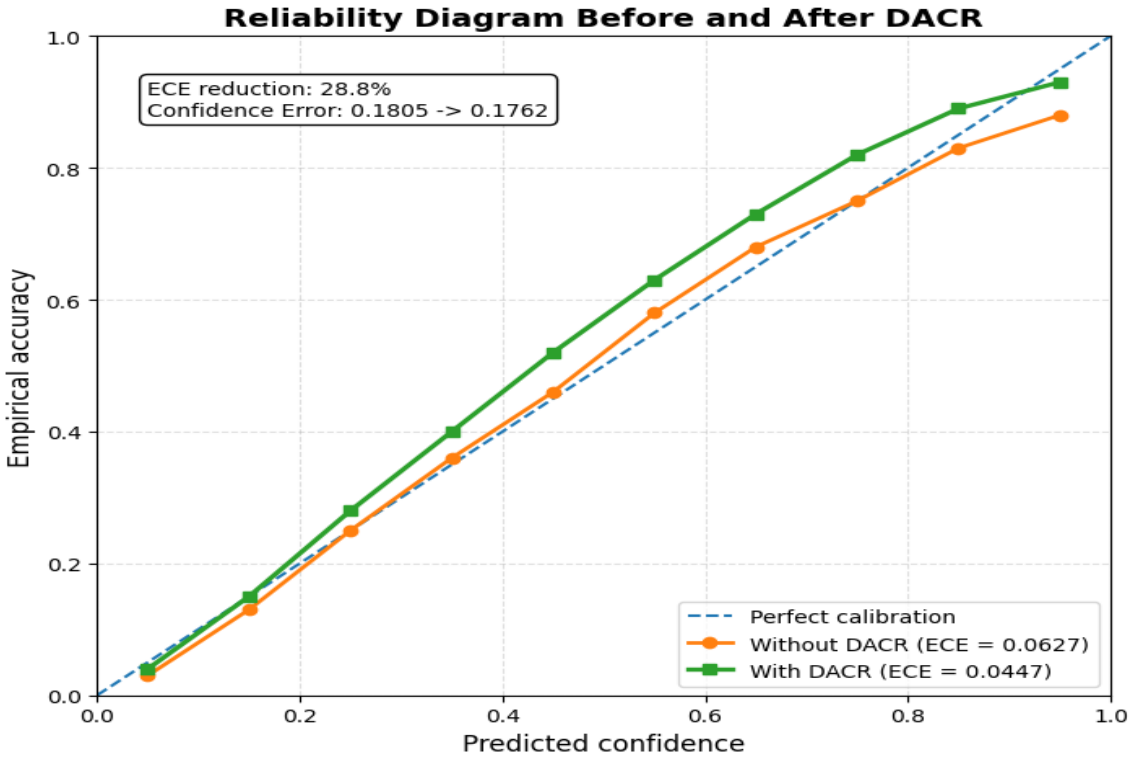


Figure 12. Reliability diagram with and without DACR calibration.

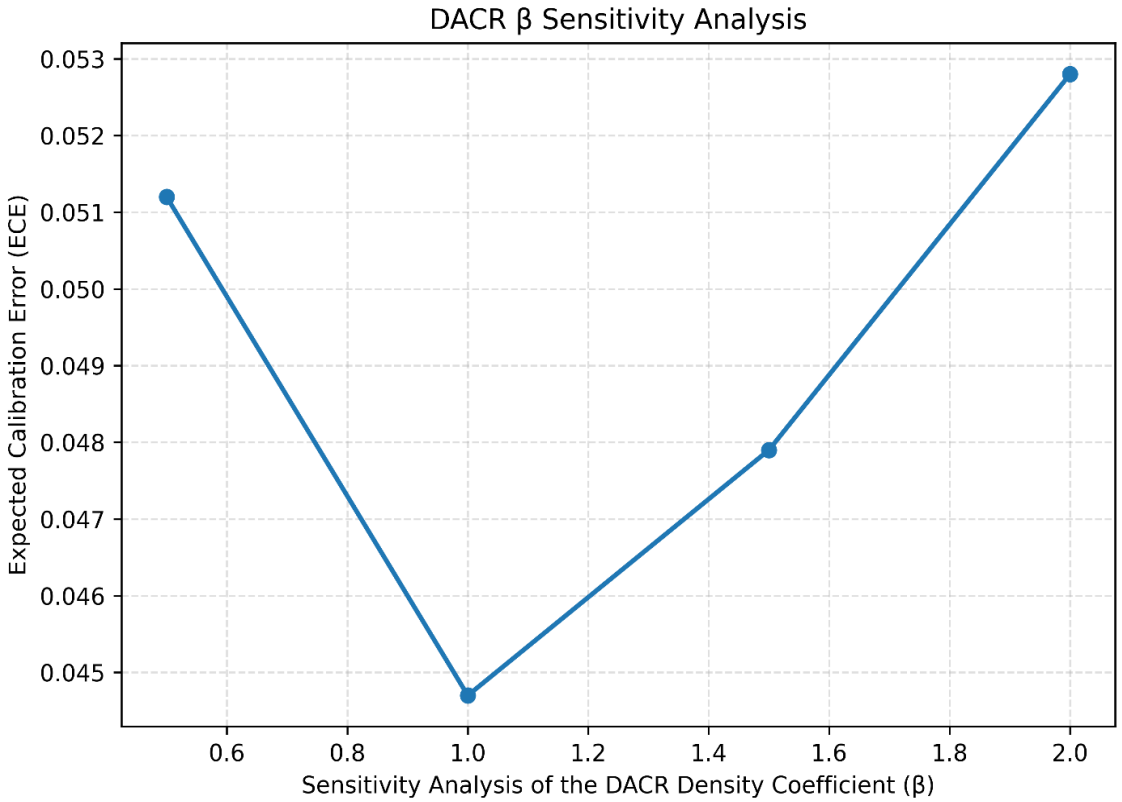


Figure 13**. Sensitivity analysis of the DACR density coefficient (β).**

Table 13. Sensitivity analysis of the DACR coefficient (β).

β	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**	**ECE**
0.5	93.05	90.88	94.82	74.80	0.0512
1.0	**93.33**	**91.21**	**95.08**	**75.45**	**0.0447**
1.5	93.10	90.95	94.90	75.02	0.0479
2.0	92.85	90.60	94.65	74.60	0.0528

Among the evaluated settings, β = 1.0 achieved the lowest ECE (0.0447) while simultaneously providing the highest detection performance, indicating the best balance between confidence calibration and localization accuracy. Accordingly, β = 1.0 was adopted for all subsequent experiments

4.7 Semi-Supervised Learning (SSL) Evaluation

The effect of annotation availability was investigated using labeled-data ratios ranging from 10% to 100%. Table 14 summarizes the corresponding detection results. Detection performance

Vol. 11, No. 06; 2026

ISSN: 2456-3676

increased with the proportion of labeled data, highlighting the importance of annotation availability for model training.

Table 14. Detection performance at different annotation ratios.

Annotation Ratio (%)	Precision (%)	Recall (%)	mAP50 (%)	mAP50-95 (%)
10	73.62504	72.91437	77.73604	34.66556
20	74.32579	75.21392	78.77141	35.23057
30	77.34508	78.38206	82.82172	38.84999
50	81.41766	79.70994	85.41755	43.10192
100	**89.78519**	**89.3401**	**93.73361**	**57.424**

Table 14 presents the detection performance obtained under different labeled-data ratios. At the 10% labeled-data setting, the model achieved 77.74% mAP@0.5 and 34.67% mAP@0.5:0.95. Increasing the labeled-data ratio to 30% improved these values to 82.82% and 38.85%, respectively. At the 50% setting, mAP@0.5 reached 85.42%, while the 100% labeled-data setting achieved 93.73% mAP@0.5 and 57.42% mAP@0.5:0.95. These results quantify the relationship between labeled-data availability and detection performance. However, the annotation ratios alone do not establish the amount of annotation time or cost saved through semi-supervised learning. A matched fully supervised baseline at each labeling ratio would be necessary to isolate the additional contribution of pseudo-label-based training.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

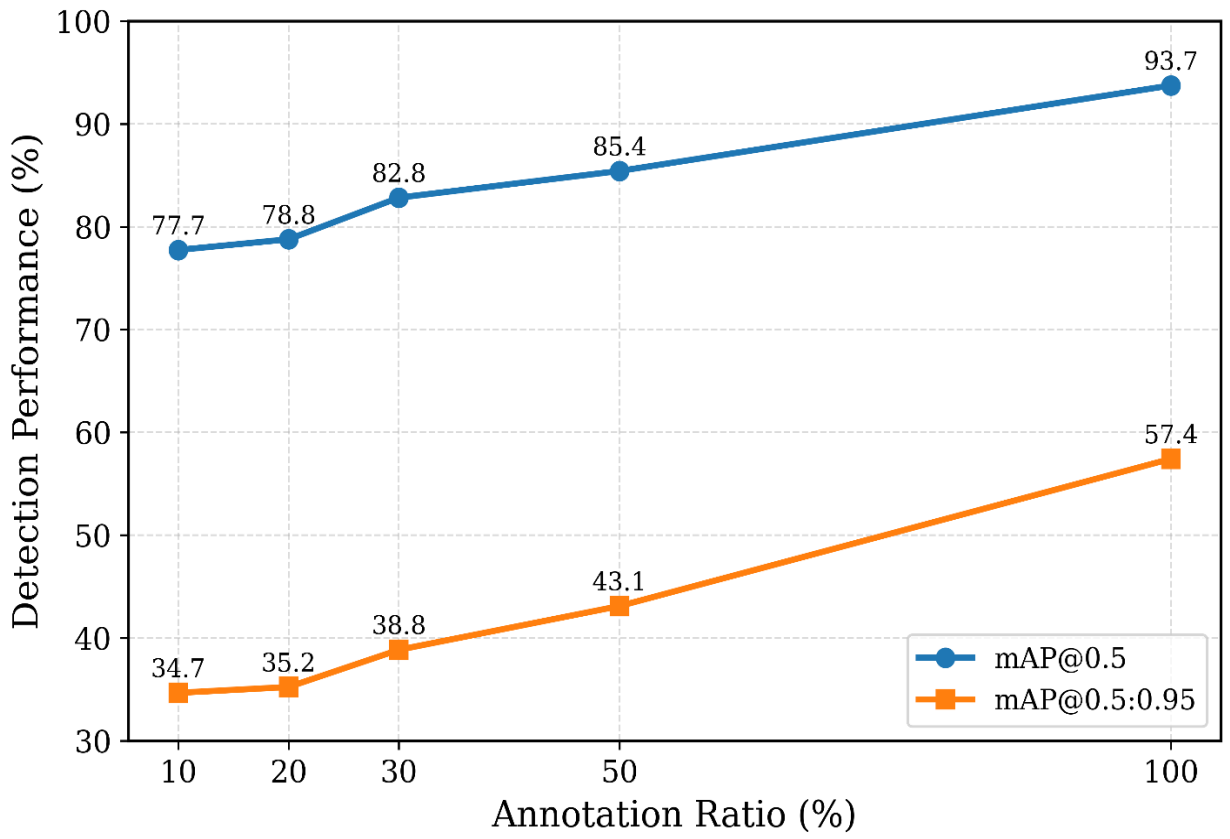


Figure 14. Detection performance under different annotation ratios.

The labeled-data-ratio experiments demonstrate the effect of annotation availability on tea-bud detection performance. The semi-supervised framework offers a mechanism for incorporating unlabeled images into training, but its actual savings in manual annotation time and cost require separate quantitative assessment.

4.8 Real-World Deployment and Application Performance

The computational efficiency of TeaBud-YOLO11++ was evaluated against representative one-stage, two-stage, and transformer-based detectors using model parameters, GFLOPs, inference speed (FPS), and detection accuracy. The results are summarized in table 15. Among the evaluated models, TeaBud-YOLO11++ achieved the highest precision (93.33%), recall (91.21%), mAP@0.5 (95.08%), and mAP@0.5:0.95 (75.45%). The proposed framework required 25.28 million parameters and 86.60 GFLOPs, achieving an inference speed of 26.72 FPS. Although YOLO11m demonstrated a higher inference speed of 32.49 FPS with lower computational complexity, TeaBud-YOLO11++ provided superior detection accuracy, particularly at stricter IoU thresholds. These results highlight the trade-off between

Vol. 11, No. 06; 2026

ISSN: 2456-3676

computational efficiency and detection performance, indicating the potential suitability of the proposed framework for intelligent tea-harvesting applications.

Table 15. Accuracy and efficiency of representative detectors.

Model	**Params (M)**	**GFLOPs**	**FPS**	**Prec. (%)**	**Rec. (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
RT-DETR-L	31.99	103.40	29.15	77.70	79.90	81.70	43.90
Faster R-CNN ResNet50-FPN	**41.30**	**133.92**	**20.18**	**91.61**	**88.97**	**87.16**	**69.36**
YOLO11m	20.11	68.50	32.49	88.83	88.09	93.14	55.83
TeaBud-YOLO11++	**25.28**	**86.60**	**26.72**	**93.33**	**91.21**	**95.08**	**75.45**

Compared with Faster R-CNN ResNet50-FPN, TeaBud-YOLO11++ improved mAP@0.5:0.95 by 6.09 percentage points while requiring fewer parameters (25.28M versus 41.30M) and lower computational complexity (86.60 versus 133.92 GFLOPs). Compared with RT-DETR-L, the proposed framework achieved a 31.55-percentage-point improvement in mAP@0.5:0.95. These comparisons illustrate its favorable accuracy–efficiency trade-off among the evaluated detectors.

Figure 15 illustrates the trade-off between detection accuracy and computational efficiency. Although YOLO11m achieved a higher inference speed, its localization accuracy was substantially lower. Faster R-CNN provided stronger localization than RT-DETR-L but required considerably greater computational resources. Overall, TeaBud-YOLO11++ achieved the most favorable balance between detection accuracy and computational efficiency among the evaluated models.

Vol. 11, No. 06; 2026

ISSN: 2456-3676

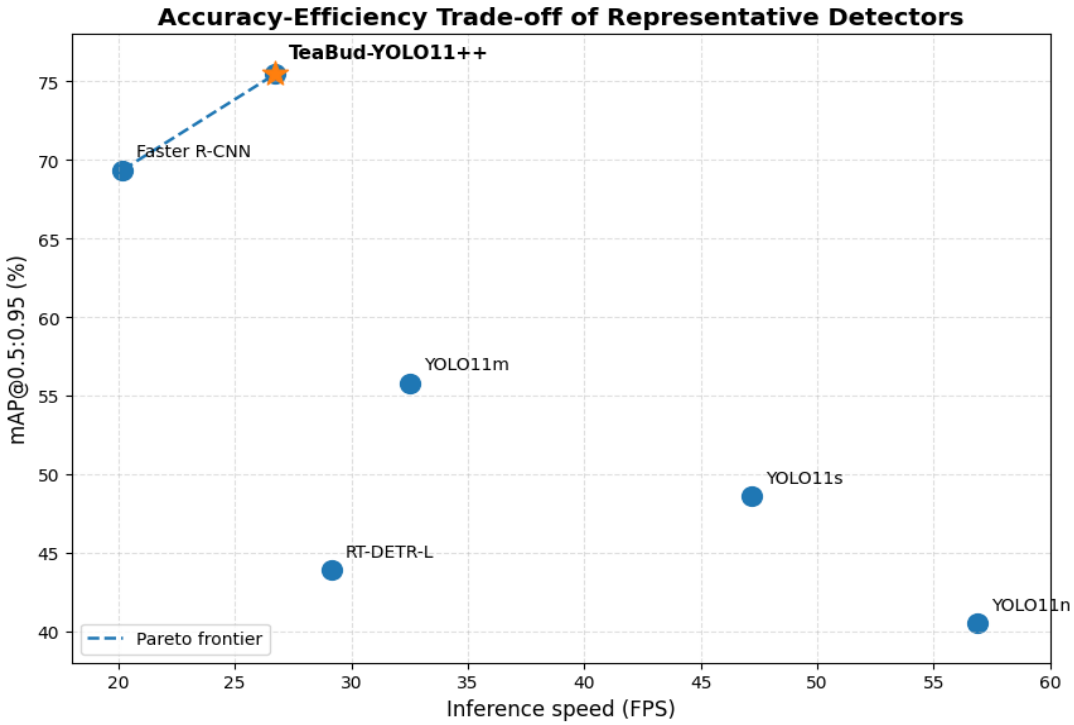


Figure 15. Accuracy–efficiency trade-off of representative detectors.

Despite the promising detection accuracy and computational performance, several limitations remain regarding real-world deployment. The reported inference speed of 26.72 FPS was obtained within the evaluated computational environment and does not establish equivalent performance on resource-constrained edge devices or embedded harvesting platforms. The model also exhibits reduced accuracy under excessive brightness and geographically different plantation conditions, indicating sensitivity to environmental variability. Furthermore, the present evaluation does not establish end-to-end harvesting performance, including robotic positioning, actuation, and operational latency. Future investigations should assess the framework on embedded hardware and integrated harvesting systems under diverse field conditions.

4.9 Comparison with State-of-the-Art Tea-Bud Detection Methods

To further assess the performance of TeaBud-YOLO11++, the proposed method was compared with representative tea-bud detection approaches reported in the literature, including YOLOv3-, YOLOv5-, and YOLOv8-based models. Because these studies were conducted using different datasets, annotation protocols, and experimental settings, the comparison is intended as a reference-level evaluation rather than a direct benchmark. Although absolute performance values are not directly comparable, this evaluation provides useful insight into the relative effectiveness of the proposed framework against existing tea-bud detection methods. Particular emphasis is

Vol. 11, No. 06; 2026

ISSN: 2456-3676

placed on annotation efficiency, detection accuracy, and practical applicability under complex field conditions. The quantitative comparison is presented in Table 16.

Table 16. Comparison with representative tea-bud detection methods.

Method	**Precision (%)**	**Recall (%)**	**mAP@0.5 (%)**	**mAP@0.5:0.95 (%)**
YOLOv3-based	74.51	69.56	72.00*	N/A
Lightweight YOLOv5	88.72	87.99	92.66	N/A
Improved YOLOv5	83.50	77.46	85.79	N/A
TBF-YOLOv8n	87.50	74.40	85.00	N/A
TeaBud-YOLO11++ (This study)	**93.33**	**91.21**	**95.08**	**75.45**

Table 16 compares TeaBud-YOLO11++ with previous tea-bud detection methods. Although several metrics are higher, differences in datasets, annotations, and evaluation settings preclude direct superiority claims. Reporting mAP@0.5:0.95 enables stricter localization assessment, suggesting promising performance under complex plantation conditions. * The original study reported mAP@0.5 only; mAP@0.5:0.95 was unavailable.

Beyond detection accuracy, practical applicability was further compared using cross-domain validation, robustness evaluation, deployment readiness, semi-supervised learning, and density-aware confidence calibration. The comparison is summarized in table 17.

Table 17. Qualitative comparison of representative tea-bud detection methods.

Method	**Backbone**	**Cross-domain Validation**	**Robustness Evaluation**	**Deployment Readiness**	**Semi-supervised Learning**	**DACR**
YOLOv3-based	YOLOv3	Limited	Limited	Partial	No	No
Lightweight YOLOv5	YOLOv5	Limited	Limited	Yes	No	No
Improved YOLOv5	YOLOv5	Limited	Limited	Yes	No	No
TBF-YOLOv8n	YOLOv8n	Limited	Partial	Yes	No	No
TeaBud-YOLO11++ (This study)	**YOLO11**	Limited	**Yes**	**Yes**	**Yes**	**Yes**

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Cross-domain validation is classified as limited because the evaluation was conducted under mixed-domain training conditions, with images from both Bangladesh and China included during training. Generalization to a completely unseen geographical domain was not established.

Representative detection examples are presented in **figure 16**, illustrating the effectiveness of TeaBud-YOLO11++ under dense, occluded, and complex field conditions.

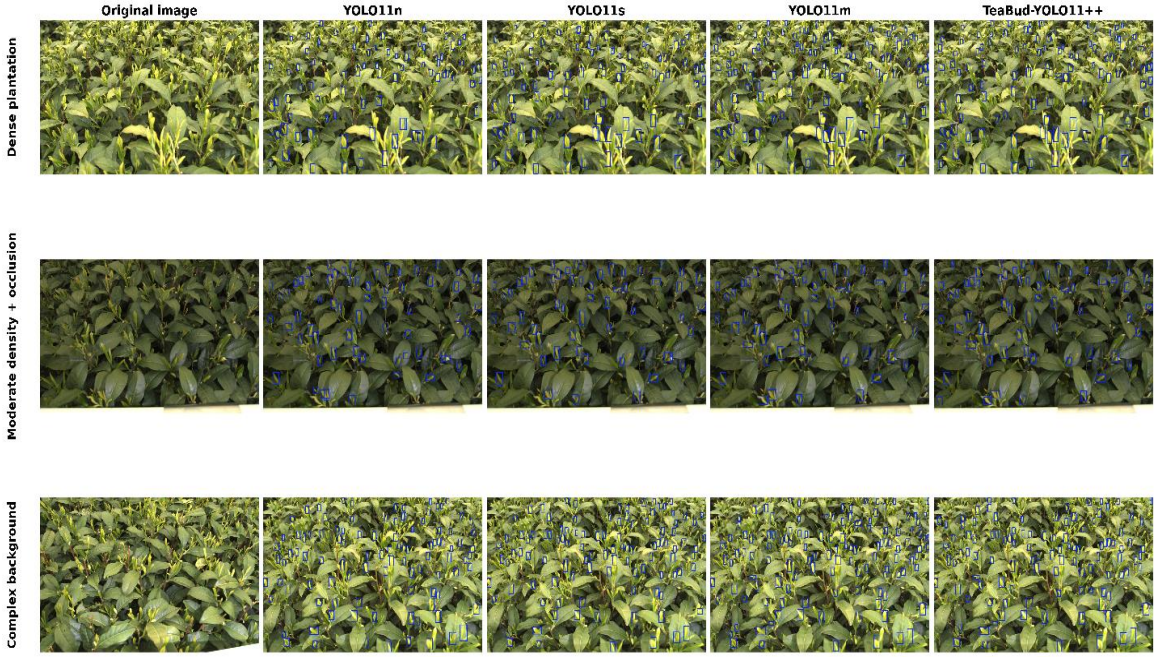


Figure 16. Qualitative Comparison of tea-bud detection results.

The comparative results indicate promising detection accuracy and confidence calibration performance for TeaBud-YOLO11++. However, comparisons with previously published tea-bud detection methods are descriptive because the studies used different datasets and evaluation protocols. Although the proposed framework demonstrates potential for intelligent harvesting applications, its deployment readiness remains to be established through integrated field and edge-device evaluations.

5. Conclusion

In this study, we proposed TeaBud-YOLO11++, an annotation-efficient semi-supervised framework for robust tea-bud detection in intelligent tea harvesting. By integrating teacher–student learning with the proposed Density-Aware Confidence Recalibration (DACR) mechanism, the framework reduced reliance on manual annotations while improving detection accuracy and confidence reliability under diverse plantation conditions. Experimental results showed that TeaBud-YOLO11++ achieved 93.33% precision, 91.21% recall, 95.08% mAP@0.5, and 75.45% mAP@0.5:0.95, outperforming representative YOLO11 variants, RT-DETR-L, and

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Faster R-CNN. Furthermore, DACR reduced the Expected Calibration Error (ECE) by 28.8%, while environmental robustness and reproducibility evaluations provided additional evidence of the framework's effectiveness.

Overall, TeaBud-YOLO11++ demonstrates promising potential for annotation-efficient tea-bud detection by combining semi-supervised learning with density-aware confidence recalibration. Nevertheless, performance variations across evaluation subsets and degradation under excessive brightness indicate opportunities for improving environmental robustness. Since images from Bangladesh and China were incorporated into the main training dataset, the present evaluation does not establish zero-shot generalization to an unseen geographical domain. Moreover, annotation-time savings and end-to-end deployment performance on embedded harvesting systems have not been quantified. Future work will focus on geographically independent and seasonal evaluations, quantitative annotation-cost assessment, and hardware-based deployment validation to further investigate the framework's practical applicability.

5.1 Limitations

While this study shows promising performance, there are some limitations. The proposed framework was validated specifically for tea-bud detection, and its applicability for other crops is to be explored. Testing the model in more regions, in different seasons, and in different imaging conditions would make the evidence for its generalizability stronger. Finally, deployment on edge devices and autonomous harvesting platforms was out of scope of this study.

5.2 Future Work

Future work will focus on evaluating TeaBud-YOLO11++ using multi-region and multi-season datasets collected under varying illumination, weather conditions, canopy densities, and growth stages. We will investigate lightweight optimization techniques, including structured pruning, quantization, and knowledge distillation, to reduce computational complexity while preserving detection accuracy. Domain adaptation methods will be explored to minimize performance degradation when transferring models between geographically distinct tea plantations. Additionally, uncertainty-aware pseudo-label selection and adaptive confidence calibration will be developed to reduce false detections in densely occluded scenes. Finally, real-world deployment on edge devices and autonomous tea-harvesting platforms will be evaluated using inference latency, frames per second (FPS), memory consumption, detection accuracy, and harvesting success rate to assess practical feasibility for precision agriculture.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFD2000600), and the National Natural Science Foundation of China (Grant Nos. U23A20175 and 32472009).

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Conflict of Interest: No competing interests.
Ethics approval and consent to participate: Not applicable.
Consent for publication: Not applicable.
Data availability: Data will be made available on request.
Materials availability: Not applicable.
Code availability: Not applicable.

Credit authorship contribution statement:
Md Morshed Alam: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Visualization, Writing – Original Draft.
Jiangsheng Gui: Supervision, Resources, Project Administration, Writing – Review & Editing.

References

Aaqil, M., Peng, C., Kamal, A., Nawaz, T., Zhang, F., & Gong, J. (2023). Tea Harvesting and Processing Techniques and Its Effect on Phytochemical Profile and Final Quality of Black Tea: A Review. *Foods*, *12*, 4467. https://doi.org/10.3390/foods12244467

Abeyrathna, R. M. R. D., Nakaguchi, V. M., Liu, Z., & Ahamed, T. (2026). Robotic Orchard and Greenhouse Fruit Harvesting Challenges and Future Trends. *IoT and AI in Agriculture: Strategies to Develop Smart Agricultural Space for In-field and In-house Crop Production*, 219–234. https://doi.org/10.1007/978-981-95-5218-4_15

Abhiram, G. (2024). The Evolution of Tea Harvesting: A Comprehensive Review of Machinery and Technological Advancements. *Journal of Biosystems Engineering*, *49*, 346–367. https://doi.org/10.1007/s42853-024-00238-9

Ao, S., Rueger, S., & Siddharthan, A. (2023). *Two Sides of Miscalibration: Identifying Over and Under-Confidence Prediction for Network Calibration*. https://doi.org/10.48550/arXiv.2308.03172

Badgujar, C. M., Poulose, A., & Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, *223*, 109090. https://doi.org/https://doi.org/10.1016/j.compag.2024.109090

Chen, C., Lu, J., Zhou, M., Yi, J., Liao, M., & Gao, Z. (2022). A YOLOv3-based computer vision system for identification of tea buds and the picking point. *Computers and Electronics in Agriculture*, *198*, 107116. https://doi.org/https://doi.org/10.1016/j.compag.2022.107116

Fahmida Islam, S., Uddin, M. S., & Bansal, J. C. (2022). Harvesting Robots for Smart Agriculture. In M. S. Uddin & J. C. Bansal (Eds.), *Computer Vision and Machine Learning in Agriculture, Volume 2* (pp. 1–13). Springer Singapore. https://doi.org/10.1007/978-981-16-9991-7_1

Guo, S., Yoon, S.-C., Li, L., Wang, W., Zhuang, H., Wei, C., Liu, Y., & Li, Y. (2023). Recognition and Positioning of Fresh Tea Buds Using YOLOv4-lighted + ICBAM Model and RGB-D Sensing. *Agriculture*, *13*(3), 518.

Hernández Hernández, G. C., Gómez Gómez, J., & Jiménez-Cabas, J. (2025). Predictive Models Based on Artificial Intelligence to Estimate Crop Yield: A Literature Review. *Agriculture*,

15(23), 2438.

Hou, G., Chen, H., Jiang, M., & Niu, R. (2023). An Overview of the Application of Machine Vision in Recognition and Localization of Fruit and Vegetable Harvesting Robots. *Agriculture*, *13*(9), 1814.

Huang, Z., Yang, D., Cheng, Z., & Jiang, X. (2025). Towards autonomous premium tea harvesting: A high-efficiency and high-quality path planning based on optimized IABFMT* algorithm in unstructured canopies. *Smart Agricultural Technology*, *12*, 101317. https://doi.org/https://doi.org/10.1016/j.atech.2025.101317

Islam, M. P., Hatou, K., Shinagawa, K., Kondo, S., Kadoya, Y., Aono, M., Kawara, T., & Matsuoka, K. (2026). Hort-YOLO: A multi-crop deep learning model with an integrated semi-automated annotation framework. *Computers and Electronics in Agriculture*, *240*, 111196. https://doi.org/https://doi.org/10.1016/j.compag.2025.111196

Kanna, K., Ramalingam, K., P, P., Ramasamy, J., & Pc, P. (2024). YOLO deep learning algorithm for object detection in agriculture: a review. *Journal of Agricultural Engineering*, *55*. https://doi.org/10.4081/jae.2024.1641

Li, Y., Zhang, Z., Zhang, J., Shi, J., Zhu, X., Chen, B., Lan, Y., Jiang, Y., Cai, W., Tan, X., Lu, Z., Peng, H., Tang, D., Zhu, Y., Tan, L., Li, K., Yang, F., & Yang, C. (2025). TeaBudNet: A Lightweight Framework for Robust Small Tea Bud Detection in Outdoor Environments via Weight-FPN and Adaptive Pruning. *Agronomy*, *15*, 1990. https://doi.org/10.3390/agronomy15081990

Liu, Z., Zhuo, L., Dong, C., Li, J., & Li, Y. (2025). TBD-Y: Automatic tea bud detection with synergistic object-spatial attention and global-local attention guided feature fusion. *Smart Agricultural Technology*, *12*, 101066. https://doi.org/https://doi.org/10.1016/j.atech.2025.101066

Nguyen, D. T., Bui, T. D., Ngo, T. M., & Ngo, U. Q. (2025). Improving YOLO-Based Plant Disease Detection Using αSILU: A Novel Activation Function for Smart Agriculture. *AgriEngineering*, *7*(9), 271.

Polpinij, J., Khoo, C. S. G., Cheng, W.-N., Khamket, T., Sibunruang, C., & Kaenampornpan, M. (2026). Reliability-Aware Multilingual Sentiment Analytics for Agricultural Market Intelligence. *Mathematics*, *14*(7), 1220.

Pugazhendi, P., Badgujar, C. M., Sapkota, R., Dhillon, R., S, R., J, J. J. S., & Ganapathy, M. R. (2026). Advances in agricultural fruit detection using you only look once (YOLO) algorithm: A review. *Smart Agricultural Technology*, *13*, 101896. https://doi.org/https://doi.org/10.1016/j.atech.2026.101896

Rahman, M. T., Bakibillah, A. S. M., Hossain, A., Ahasan, A., Basher, M. N., Oyshe, K. U., & Mariam, A. (2026). The Role of AI-Integrated Drone Systems in Agricultural Productivity and Sustainable Pest Management. *AgriEngineering*, *8*(4), 142.

Sapna, Gauttam, H., Chauhan, V., Pattanaik, K. K., Trivedi, A., & Ghosh, H. (2026). A comprehensive review of Edge Computing empowered smart agriculture: Trends, Opportunities and Future directions. *Computers and Electronics in Agriculture*, *241*, 111252. https://doi.org/https://doi.org/10.1016/j.compag.2025.111252

Sbai, Z. (2025). Deep Learning Models and Their Ensembles for Robust Agricultural Yield

Prediction in Saudi Arabia. *Sustainability*, *17*, 5807. https://doi.org/10.3390/su17135807

Shi, J., Zheng, Y., Wang, R., & Li, Y. (2025, 26–28 Sept. 2025). TEA-YOLO: A Detection Model for Tea Buds Based on improved YOLOv11 and Soft Actor-Critic. 2025 3rd International Conference on Algorithm, Image Processing and Machine Vision (AIPMV),

Singh, P., Sharma, H., & Kuntal, P. (2020). Agricultural Robots for Harvesting and Planting. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *11*, 1112–1118. https://doi.org/10.61841/turcomat.v11i1.14410

Soleimani, A., Abbaspour-Fard, M. H., Sapkota, R., & Sanaeifar, A. (2026). Artificial intelligence in agricultural machinery & robotics: A systematic review and future outlook. *Applied Soft Computing*, *194*, 114889. https://doi.org/https://doi.org/10.1016/j.asoc.2026.114889

Song, X., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. (2025a). Agricultural Image Processing: Challenges, Advances, and Future Trends. *Applied Sciences*, *15*(16), 9206.

Song, X., Yan, L., Liu, S., Gao, T., Han, L., Jiang, X., Jin, H., & Zhu, Y. (2025b). Agricultural Image Processing: Challenges, Advances, and Future Trends. *Applied Sciences*, *15*, 9206. https://doi.org/10.3390/app15169206

Song, Y., Zheng, Z., Zhang, H., Liu, Z., Chen, L., Ning, J., & Dai, Q. (2025). High-Precision Method for Simultaneous Tea Bud and Picking Point Detection Using Morphology Heatmap Labels. *Applied Engineering in Agriculture*, *41*, 627–641. https://doi.org/10.13031/aea.16407

Wang, J., Guo, Y., Han, Y., Zhu, Y., & Lan, Y. (2026). A top-view transformer-based framework for tiny tea bud detection with gradient-guided attention and position-sensitive loss. *Computers and Electronics in Agriculture*, *240*, 111123. https://doi.org/https://doi.org/10.1016/j.compag.2025.111123

Wang, W., Li, C., Xi, Y., Gu, J., Zhang, X., Zhou, M., & Peng, Y. (2025). Research Progress and Development Trend of Visual Detection Methods for Selective Fruit Harvesting Robots. *Agronomy*, *15*(8), 1926.

Wang, W., Xi, Y., Gu, J., Yang, Q., Pan, Z., Zhang, X., Xu, G., & Zhou, M. (2025). YOLOv8-TEA: Recognition Method of Tender Shoots of Tea Based on Instance Segmentation Algorithm. *Agronomy*, *15*(6), 1318. https://doi.org/10.3390/agronomy15061318

Wang, Y., Lu, J., Wang, Q., & Gao, Z. (2024). A method of identification and localization of tea buds based on lightweight improved YOLOV5 [Original Research]. *Frontiers in Plant Science*, *Volume 15 - 2024*. https://doi.org/10.3389/fpls.2024.1488185

Wu, T., Zhou, L., Zhao, Y., Qi, H., Pu, Y., Zhang, C., & Liu, Y. (2025). Applications of deep learning in tea quality monitoring: a review. *Artificial Intelligence Review*, *58*. https://doi.org/10.1007/s10462-025-11335-2

Xie, S., & Sun, H. (2023). Tea-YOLOv8s: A Tea Bud Detection Model Based on Deep Learning and Computer Vision. *Sensors*, *23*, 6576. https://doi.org/10.3390/s23146576

Yang, J., Li, X., Wang, X., Fu, L., & Li, S. (2024). Vision-Based Localization Method for Picking Points in Tea-Harvesting Robots. *Sensors*, *24*(21), 6777.

Yang, Z., Ma, W., Lu, J., Zhiwei, T., & Peng, K. (2023). The Application Status and Trends of

Vol. 11, No. 06; 2026

ISSN: 2456-3676

Machine Vision in Tea Production. *Applied Sciences*, *13*, 10744. https://doi.org/10.3390/app131910744

Yildiz, M., Mert, M. A. A., & Mansoor, S. (2026). From classical practices to precision agriculture: a multidisciplinary review of tea (Camellia sinensis) [Review]. *Frontiers in Plant Science*, *Volume 17 - 2026*. https://doi.org/10.3389/fpls.2026.1848455

Yu, C., Xue, Y., Zhang, L., An, X., Liu, C., & Chen, L. (2026). YOLO-MEST: A re-parameterized multi-scale fusion model with enhanced detection head for high-accuracy tea bud detection. *Information Processing in Agriculture*, *13*(2), 163–175. https://doi.org/https://doi.org/10.1016/j.inpa.2025.10.001

Yu, R., Xie, Y., Li, Q., Guo, Z., Dai, Y., Fang, Z., & Li, J. (2024). Development and Experiment of Adaptive Oolong Tea Harvesting Robot Based on Visual Localization. *Agriculture*, *14*(12), 2213.

Zhang, J., Obrien, E., Kong, X., & Deng, L. (2023). Factors affecting the accuracy of a computer vision-based vehicle weight measurement system. *Measurement*, *224*, 113840. https://doi.org/10.1016/j.measurement.2023.113840

Zhang, T., Yang, Q., Tong, X., Hu, L., & Shao, J. (2026). A transformer-based path fusion detector network for tea bud detection. *Smart Agricultural Technology*, *14*, 102321. https://doi.org/https://doi.org/10.1016/j.atech.2026.102321

Zhang, X., Ye, S., Wang, Y., Gouda, M., Hu, Y., Islam, M., He, Y., & Li, X. (2025). Intelligent Transformation of the Tea Industry: Artificial intelligence applications in management, harvesting, and processing. *Trends in Food Science & Technology*, 105510. https://doi.org/10.1016/j.tifs.2025.105510

Zhao, S., Ye, C., Lian, C., Mei, L., Wu, L., & Chen, J. (2026). Fine-Grained Detection and Sorting of Fresh Tea Leaves Using an Enhanced YOLOv12 Framework. *Foods*, *15*(3), 544.

Zhu, L., Zhang, Z., Lin, G., Chen, P., Li, X., & Zhang, S. (2023). Detection and Localization of Tea Bud Based on Improved YOLOv5s and 3D Point Cloud Processing. *Agronomy*, *13*(9), 2412.