Towards Smart Agriculture: A Deep Learning and Explainable AI Framework for Automated Rice Panicle Detection in Bangladesh

Authors
Raiyan Gani (2025-2-96-012)

Email:
raiyan.gani02@gmail.com

ABSTRACT

Rice panicle detection is a crucial step in yield estimation, yet traditional manual methods remain labor-intensive and error-prone. This study presents a deep learning–based framework for automated panicle detection in Bangladeshi rice fields, integrating transfer learning, self-supervised pre-training, and explainable AI. A dataset of 2,193 high-resolution drone images was collected in Gazipur, followed by preprocessing and augmentation. We evaluated state-of-the-art models, including YOLOv8, YOLOv9, and DINO-YOLO variants, with Grad-CAM for interpretability and Weighted Boxes Fusion (WBF) for prediction refinement. Results show that YOLOv9 achieved the best performance with an mAP@0.50 of 0.770, surpassing YOLOv8 (0.759). Both models maintained high precision (>0.68) and recall (>0.71), confirming strong detection reliability in field conditions. Grad-CAM visualizations demonstrated that the models consistently focused on panicle regions, increasing interpretability and trust. While DINO-YOLO models underperformed in accuracy, they highlighted the potential of self-supervised learning to reduce annotation requirements. Overall, the proposed study delivers accurate and interpretable rice panicle detection, supporting scalable, automated yield monitoring. This contributes directly to smart agriculture in Bangladesh, reducing manual effort and enhancing food security.

Keywords - Rice Panicle Detection, Deep Learning, Self-Supervised Learning (SSL), Explainable AI (XAI), Precision Agriculture

INTRODUCTION

Agriculture is undergoing a technological revolution, driven by the integration of artificial intelligence (AI), machine learning, and advanced sensing technologies. AI offers new opportunities to enhance agricultural productivity, improve sustainability, and mitigate the challenges of climate change and resource scarcity. Hamed et al. [1] emphasized that AI in agriculture not only contributes to higher yields but also supports sustainable practices, enabling smarter decisions in crop management and food production. These advancements are particularly significant in developing nations such as Bangladesh, where agriculture is the backbone of the economy and food security. Rice, being the staple crop in Bangladesh, plays a central role in ensuring national food supply and supporting rural livelihoods. However, traditional methods for estimating rice yield, such as manual panicle counting and farmer surveys, are labor-intensive, time-consuming, and error-prone. These limitations hinder the ability of policymakers and farmers to obtain timely, accurate insights into crop conditions. To overcome these barriers, AI-powered approaches are increasingly being adopted for crop monitoring, yield estimation, and disease detection. More importantly, computer vision and deep learning have transformed agricultural monitoring by enabling accurate organ-level crop detection and classification. In computer vision, object detection frameworks such as *EfficientDet* [2] have pioneered scalable and efficient detection methods that achieve a balance between accuracy and computational efficiency. Agricultural researchers have adapted these frameworks for practical applications. For example, Wang et al. developed a deep learning pipeline for rice panicle detection and counting in field conditions [3]. Xiang et al. introduced *YOLO Pod*, a fast and accurate multi-task model for dense soybean pod counting [4]. Similarly, Sun et al. proposed *WDN*, a one-stage detection network designed for wheat heads with high precision [5]. Koirala et al. applied deep learning to classify mango panicle stages, proving the adaptability of convolutional neural networks (CNNs) to agricultural crop monitoring [6]. Together, these studies demonstrate the potential of deep learning-based detection frameworks for improving yield estimation, resource allocation, and decision-making in agriculture.

However, the successful deployment of AI in agriculture is constrained by two key factors: (i) the lack of large annotated datasets and (ii) the limited generalization of models trained in one environment to new crops or regions. Annotating large volumes of agricultural images—such as identifying rice panicles in drone or satellite imagery—is both time-intensive and costly. Consequently, supervised learning approaches often struggle when applied to datasets from different environments, as variations in lighting, soil conditions, and crop morphology reduce performance. To mitigate these challenges, self-supervised learning (SSL) has emerged as a promising paradigm. Unlike traditional supervised learning, SSL leverages unlabeled data to learn robust feature representations, requiring only minimal annotation for fine-tuning. This is particularly valuable in agriculture, where labeled datasets are scarce but large volumes of raw imagery are available through drones and satellites. Recent research highlights the promise of SSL in agricultural contexts. Yilma et al. proposed *Attentive Self-supervised Contrastive Learning (ASCL)* for plant disease classification, achieving superior accuracy under limited-label conditions [8]. Mamun et al. conducted a systematic review, underlining SSL's role in reducing annotation costs and improving generalization [9]. Wang et al. extended SSL to transformer-based pre-training using the *General Plant Infection* dataset, enabling cross-domain adaptability in plant health monitoring [10].

Moreover, SSL integrates effectively with transfer learning, where pre-trained models are fine-tuned for specific agricultural tasks. While transfer learning has already been used for panicle and organ detection [3–6], combining it with SSL enhances robustness and reduces reliance on annotated data. In this study, pre-trained models are further augmented with explainable AI (XAI) techniques to ensure interpretability. Applying XAI enables researchers and agricultural stakeholders to visualize which parts of an image influence the model's decisions. For example, heatmaps and feature attribution methods can identify whether the model focuses correctly on panicle regions rather than irrelevant background features. This transparency is critical for building trust in AI applications, particularly in high-stakes domains like agriculture, where interpretability supports both scientific validation and farmer adoption. Parallel to these developments, automation in agriculture is advancing rapidly. Reviews on self-driving and automatic navigation technologies highlight how AI can transform agricultural machinery into intelligent, autonomous systems [8]. When coupled with SSL, transfer learning, and XAI, such automation has the potential to create a new generation of precision agriculture tools—scalable, interpretable, and adaptable across different crops and regions. Bangladesh presents both a challenge and an opportunity in this context. As a rice-dependent country, ensuring timely and accurate yield estimation is critical to food security. Yet, its agricultural sector faces complex issues such as climate variability, floods, and droughts, which complicate prediction models. Remote sensing technologies, including high-resolution satellite imagery and UAV-based monitoring, provide a way to capture large-scale crop information across the country. When paired with fine-tuned AI models, these images can reveal crop conditions, detect panicles, and estimate yields far more accurately than traditional methods. However, most existing AI frameworks have been trained and validated in regions outside Bangladesh, limiting their direct applicability. Developing a framework that integrates remote sensing, SSL-based pre-training, transfer learning, and XAI offers a promising pathway to overcome these limitations.

The motivation for this study arises from the convergence of these needs and opportunities. Previous research in object detection [2–6] has demonstrated the effectiveness of deep learning for agricultural monitoring. Reviews on agricultural automation [8] point toward an AI-driven future in farming. SSL-based methods [9–11] highlight a transformative solution for data scarcity, while AI in agriculture more broadly has been shown to enhance sustainability and productivity [1]. Yet, there is little work that applies SSL and transfer learning to rice panicle detection in Bangladesh, while also incorporating XAI for interpretability.

Therefore, this paper proposes *a framework for rice panicle detection in Bangladesh using remote sensing and fine-tuned transfer learning*. Specifically, the framework leverages self-supervised pre-training on large-scale unlabeled agricultural imagery, followed by transfer learning fine-tuning on smaller annotated datasets for panicle detection. Furthermore, XAI techniques are applied to the pre-trained models, ensuring transparency and interpretability of detection results. The main objectives of this study are:

1. To design a panicle detection pipeline integrating remote sensing data, SSL, and fine-tuned transfer learning.

2. To apply XAI methods to explain model predictions, improving interpretability and trust in the framework.

3. To evaluate the potential of the framework for scalable rice yield estimation in Bangladesh.

By addressing these objectives, this research contributes to advancing AI-driven agriculture in multiple ways. From a scientific perspective, it extends the integration of SSL, transfer learning, and XAI into agricultural detection tasks. From a practical perspective, it provides a locally adapted, interpretable, and scalable framework for rice monitoring in Bangladesh, directly supporting national food security and the vision of sustainable, smart agriculture.

Literature Review

Artificial intelligence has been increasingly applied in agriculture, particularly in rice research, to address challenges in disease detection, phenotyping, yield estimation, and growth monitoring. Recent studies have explored advanced imaging technologies such as X-ray, infrared, and UAV-based imagery combined with deep learning models for accurate and efficient analysis. These works provide a strong foundation for rice panicle detection and information management, which is the primary focus of this study. Yang et al. [11] introduced an innovative rice blast detection approach by fusing 4D light field refocusing with depth information, enabling the system to capture multi-dimensional plant features and achieve higher accuracy in identifying early-stage blast infections. This method demonstrates how advanced imaging fusion can strengthen disease detection in rice, providing more reliable support for precision agriculture.

Similarly, Yu et al. [12] designed an integrated rice panicle phenotyping framework that combines X-ray and RGB scanning with deep learning algorithms, achieving accurate and efficient trait extraction. Their multi-modal strategy significantly improved panicle structural analysis, highlighting the role of combined imaging and AI in agricultural trait phenotyping. Hu et al. [13] developed a nondestructive 3D image analysis pipeline using X-ray computed tomography (CT), allowing precise measurement of rice grain traits without destroying samples. This work contributes to detailed grain morphology studies, supporting breeding and yield analysis with high-throughput phenotyping tools. In terms of growth stage monitoring, Desai et al. [14] proposed a deep learning-based system for automatic estimation of heading dates in paddy rice using field-acquired image datasets. Their approach reduced manual labor and provided scalable solutions for growth stage prediction. Deng et al. [15]

advanced this area further by presenting a CNN-based framework to automatically detect productive tillers in rice plants, a key factor influencing yield estimation. Their method offers a robust tool for tiller identification, supporting accurate agronomic assessments. For yield trait analysis, Sun et al. [16] introduced LED transmission imaging in combination with cloud computing to provide intelligent and scalable solutions for rice yield trait monitoring. This framework enables large-scale deployment, increasing the efficiency of trait analysis in agricultural fields.

Guo et al. [17] presented an automatic deep learning approach for rice seed setting rate calculation using image segmentation and recognition. Their system showed high accuracy in distinguishing fertile and sterile spikelets, reducing the effort of manual counting. Singh [18] applied a YOLO-based model called YOLOmaturity to estimate paddy grain maturity, providing a fast and real-time method for determining harvest readiness. Building on advanced sensing, Zhou et al. [19] developed a thermal infrared de-ghosting and multi-level segmentation algorithm for detecting rice seed setting rates, proving that thermal imagery can complement RGB data to enhance detection robustness under variable field conditions.

Meanwhile, UAV-based approaches have gained popularity in rice monitoring. Lyu et al. [20] combined UAV time-series imagery with novel machine learning techniques to estimate rice heading dates for breeding purposes, offering scalable monitoring for large fields. Barreto et al. [21] demonstrated the effectiveness of UAV-based seedling counting, initially for sugar beet and later extended to maize and strawberry, indicating that UAV methods can be adapted across multiple crops for early-stage monitoring. Focusing specifically on panicle measurement, Ziyue et al. [22] introduced Panicle Ratio Network (PRNet), leveraging ultra-high-definition aerial imagery and deep learning to streamline the detection and measurement of rice panicles at scale. Similarly, Deng et al. [23] applied CNN models to estimate grain number automatically, contributing to more precise yield predictions by reducing dependence on manual counting. Zhou et al. [24] presented a UAV-based panicle counting method powered by deep learning, which enabled accurate and efficient panicle detection across large rice fields. Their work supports the automation of yield estimation in real-world agricultural environments.

Guo et al. [25] used time-series RGB images to monitor flowering dynamics in rice, helping identify flowering patterns under field conditions, which is crucial for crop management and breeding. On the other hand, Liu et al. [26] proposed an improved lightweight YOLOv5s model for growth period recognition in rice, achieving real-time performance with reduced computational cost, making it suitable for deployment in resource-limited field settings. Jocher et al. [27] released YOLOv5, a state-of-the-art real-time detection model widely applied in rice panicle and disease detection research, offering an effective baseline for precision agriculture.

Bochkovskiy et al. [28] further advanced this direction with YOLOv7, introducing new training strategies and architectural improvements to enhance detection accuracy. Building on this, Jocher et al. [29] developed YOLOv8, which supports both segmentation and detection, providing versatile functionality for agricultural datasets. Finally, Solovyev and Wang [30] proposed the Weighted Boxes Fusion (WBF) technique, which ensembles multiple object detection outputs to refine bounding box predictions, thus improving detection reliability in challenging agricultural environments. Recent advancements in artificial intelligence and self-supervised learning have played a crucial role in agricultural automation, particularly in plant reconstruction, disease detection, and rice phenotyping. For instance, Ci et al. [31] introduced SSL-NBV, a self-supervised-learning-based next-best-view algorithm that enhances robotic efficiency in 3D plant reconstruction, contributing significantly to precision agriculture.

Similarly, Cao et al. [32] addressed the challenge of small-sample cucumber disease detection through multimodal self-supervised learning, which demonstrates the potential of reducing dependency on large labeled datasets. Building on this, Wang et al. [33,35] applied self-supervised learning approaches for classifying plant leaf diseases, showcasing improved generalization and robustness across varying crop conditions. In addition, Al Mamun et al. [34] provided a systematic review of plant disease detection methods using self-supervised learning, further highlighting its transformative role in overcoming data scarcity in agriculture. Complementing these approaches, Zhao et al. [36] proposed a contrastive learning framework with domain adaptation for leaf disease identification, enabling models to adapt more effectively across different environments. Beyond plant disease identification, research in rice phenotyping has also made significant progress.

Tan et al. [37] developed RiceRes2Net, a deep learning-based framework for detecting in-field rice panicles and recognizing growth stages, ensuring accurate monitoring throughout the crop lifecycle. In grain-level studies, Deng et al. [38] applied deep learning for automated grain counting on rice panicles, which provides precise yield estimation at a micro level. Likewise, Wei et al. [39] utilized UAV-based imagery and deep convolutional networks to generate rice density prescription maps during the ripening stage, which facilitates site-specific crop management. Earlier, Zhou et al. [40] pioneered UAV-based rice panicle counting using deep learning, paving the way for large-scale automated yield assessment.

METHODOLOGY

This study proposes an end-to-end framework for automated rice panicle detection, integrating drone-based image acquisition, systematic preprocessing and augmentation, state-of-the-art object detection models, and explainable AI for interpretability. The methodology is structured into five components: dataset acquisition, data preprocessing, model selection, model training, and interpretability analysis.

Dataset Description

In this study on rice panicle detection, data was collected using high-quality drones in Gazipur City, located near Dhaka, Bangladesh [41]. The aim was to capture precise and region-specific data relevant to rice panicles. A skilled and experienced drone operator managed the data collection process, ensuring the acquisition of accurate and high-resolution images of rice panicles and their surrounding environments. The operator's expertise played a crucial role in maneuvering the drones effectively, capturing diverse visual data from the region.

An overview of the data collection process is illustrated in Fig. 1. After the data was collected, the next phase involved comprehensive data preprocessing to prepare the dataset for model training. This included image annotation, enhancement, and augmentation, with careful attention given to the consistency and quality of the images. The process also ensured clarity, uniformity, and proper labelling of images, with annotations made where necessary to guarantee the dataset's suitability for training purposes.

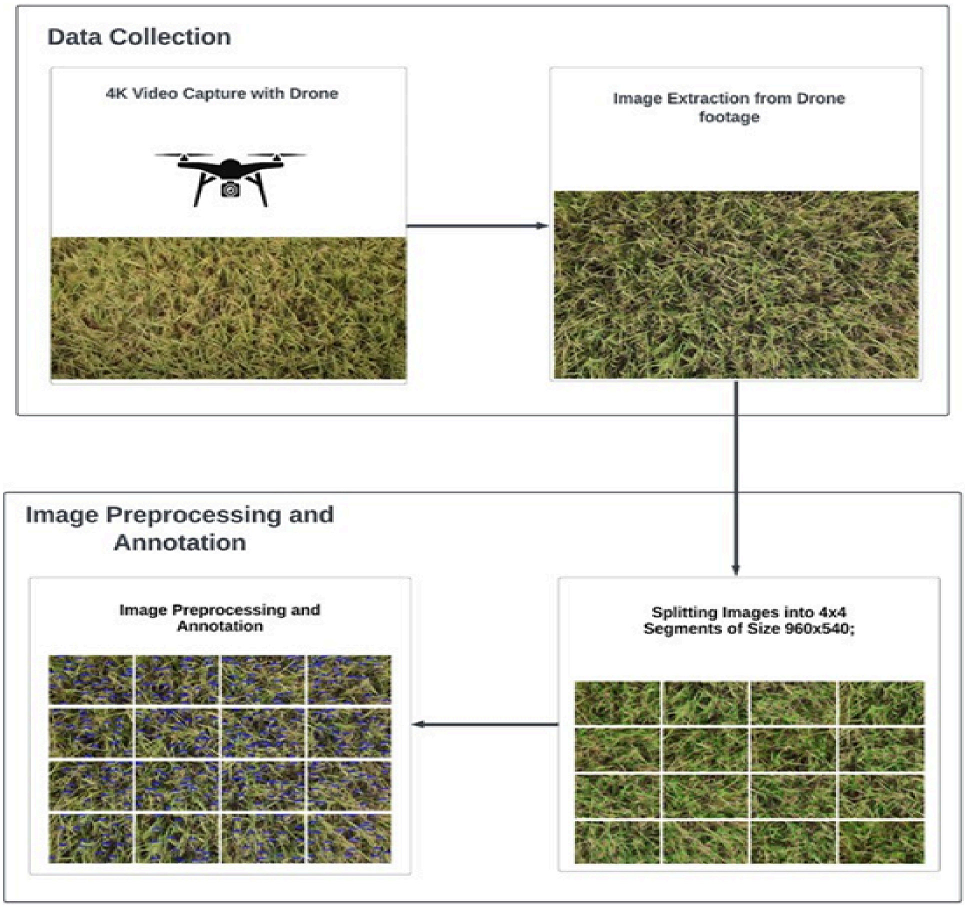

Fig. 1. Data collection Overview

Data Augmentation and Pre-processing

Following collection, all images underwent annotation through a semi-supervised pipeline to label panicle regions. To enhance model generalization, extensive augmentation techniques were applied, including:

- Horizontal and vertical flipping
- Rotations at 90° intervals
- Adjustments of hue, saturation, and brightness (±25°)
- Exposure correction (–5% to +5%)
- Mosaic transformations

These operations generated a diverse training set, improving robustness against variations in illumination, background, and orientation. Representative augmented samples are shown in Fig. 2.

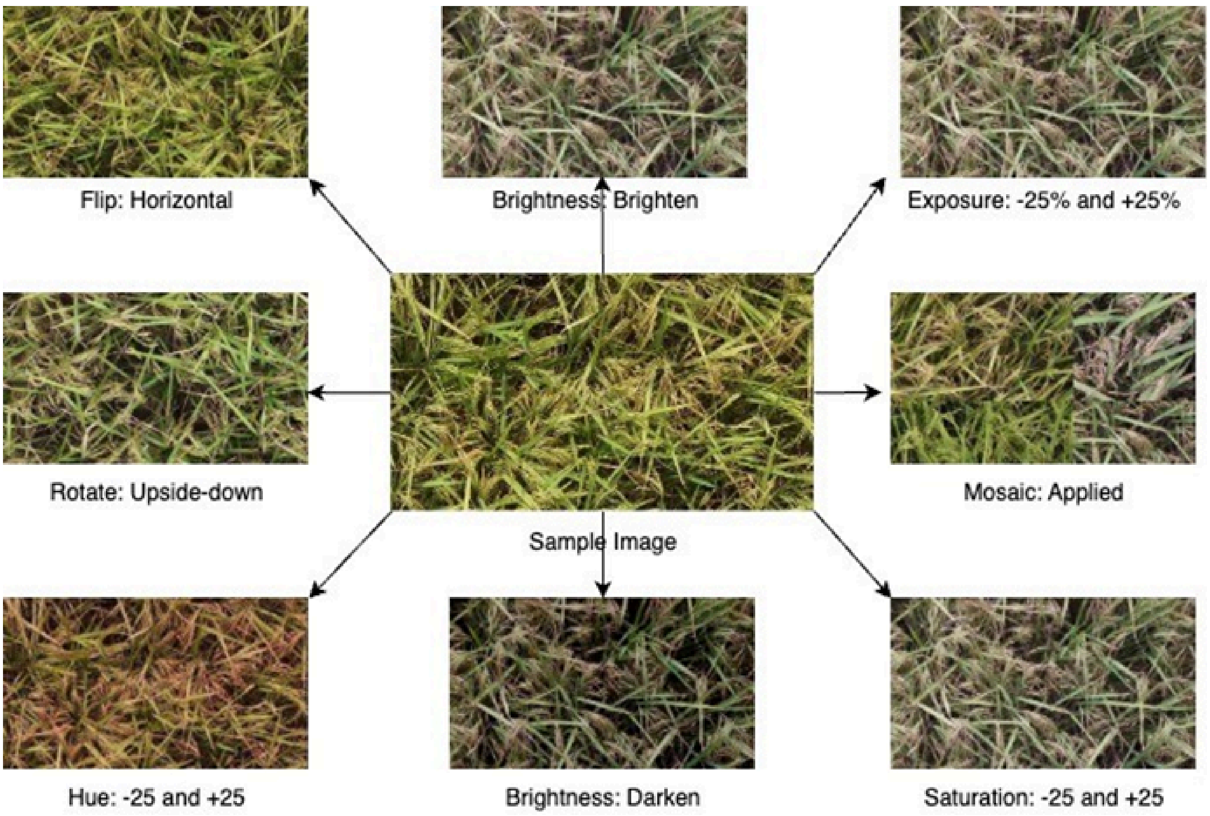

Fig. 2. Data Augmentation Samples

Dataset Description

In this study, we employ cutting-edge deep learning techniques to address the critical challenge of rice panicle detection, a key task in enhancing rice production efficiency. Our focus is on evaluating the performance of advanced deep learning models, including YOLOv8 with Grad-CAM, YOLOv9 with Grad-CAM, YOLOv10 with Grad-CAM, DINO-YOLOv8, DINO-YOLOv9, and DINO-YOLOv10. These models are chosen for their exceptional performance in object detection, which is essential for accurate agricultural interventions.

To improve the overall prediction quality, we apply Weighted Boxes Fusion (WBF), which merges the predictions from these different models. This technique helps refine the accuracy of the combined outputs, ensuring more reliable results. The experimental setup is carefully designed, with custom adjustments made to image sizes, training epochs, and augmentation methods. This approach ensures an optimal balance between computational efficiency and model performance.

Furthermore, we employ transfer learning, utilizing pre-trained models to speed up the training process and enhance the generalization capability for rice panicle detection. Our comprehensive methodology aims to provide an automated and highly accurate solution for rice panicle detection, advancing the field of precision agriculture. This solution is particularly important for improving rice cultivation practices and ensuring food security in regions such as Bangladesh.

Applied Models

For efficient and accurate rice panicle detection, we meticulously evaluated several state-of-the-art object detection models, ultimately selecting YOLOv8 with Grad-CAM, YOLOv9 with Grad-CAM, DINO-YOLOv8, DINO-YOLOv9, and DINO-YOLOv10 for their outstanding performance across various key metrics. This selection of models ensures a robust and versatile approach to rice panicle detection, with a focus on accuracy, speed, and the ability to generalize effectively across different detection challenges.

YOLO MODELS

The YOLO series, including YOLOv8 and YOLOv9, is renowned for its real-time object detection capabilities, providing an ideal balance between speed and accuracy. These models achieved high validation and testing accuracies, as demonstrated in our evaluation results, making them particularly well-suited for rice panicle detection.

YOLOv8 Architecture

YOLOv8, the latest version of the YOLO series, introduces several improvements to enhance both accuracy and efficiency. It uses a CSPDarknet backbone with C2f modules for stronger feature extraction while keeping the model lightweight. The neck combines Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) layers to fuse information from different scales, allowing better detection of small and overlapping rice panicles. Unlike earlier versions, YOLOv8 adopts an anchor-free detection head, which separately handles classification and localization, improving bounding box precision. These modifications make YOLOv8 faster to train, more accurate, and well-suited for real-time rice panicle detection.

YOLOv9 Architecture

YOLOv9 builds on the strengths of earlier YOLO versions with further improvements in accuracy and generalization. It introduces an advanced backbone with programmable gradient information (PGI), which helps the network preserve richer features during training. Similar to YOLOv8, it employs multi-scale feature fusion, but with optimized pathways that improve detection of small and dense objects such as rice panicles. YOLOv9 also refines the anchor-free detection head, enhancing both localization precision and classification confidence. These innovations make YOLOv9 more robust in challenging field conditions, offering higher detection performance while maintaining real-time efficiency.

DINO-YOLO MODELS

Additionally, the DINO-YOLO series such as DINO-YOLOv8, DINO-YOLOv9, and DINO-YOLOv10 was incorporated due to its promising performance in self-supervised learning and its ability to leverage pre-trained models, enhancing the overall detection capabilities.

DINO-YOLOv8 Architecture

DINO-YOLOv8 combines the YOLOv8 detection framework with self-supervised pre-training using the DINO approach (Distillation with No Labels). The model first learns feature representations from large amounts of unlabeled images, allowing it to capture general patterns before fine-tuning on annotated rice panicle data. By integrating DINO embeddings into the YOLOv8 backbone, the network gains stronger semantic understanding and improved robustness under limited training samples. The anchor-free head from YOLOv8 is retained, ensuring efficient localization and classification, while the self-supervised pre-training enhances generalization to field variations such as lighting, occlusion, and background noise. This combination makes DINO-YOLOv8 particularly valuable for agricultural tasks where annotated data are scarce..

DINO-YOLOv9 Architecture

DINO-YOLOv9 integrates the improved YOLOv9 detection pipeline with self-supervised pre-training using the DINO framework. The model benefits from YOLOv9’s advanced backbone with programmable gradient information (PGI) while leveraging DINO embeddings to strengthen feature representation from unlabeled agricultural images. This combination allows the model to adapt better to small, dense, and complex targets such as rice panicles. The anchor-free detection head ensures efficient classification and localization, while the self-supervised initialization improves generalization when annotated data are limited. As a result, DINO-YOLOv9 offers a balance between high detection accuracy and reduced dependency on large-scale labeled datasets, making it suitable for real-world agricultural scenarios..

Explainable AI

To ensure transparency in decision-making, Grad-CAM (Gradient-weighted Class Activation Mapping) was applied to the YOLO models. This technique highlights the image regions that most influenced predictions, enabling validation of whether the models attended correctly to panicle clusters. Such interpretability is essential for building trust in agricultural AI applications, particularly for deployment among farmers and policymakers.

RESULTS AND ANALYSIS

This section presents the experimental outcomes of the proposed rice panicle detection framework, evaluating both YOLO-based and self-supervised DINO-YOLO models. The performance analysis considers key object detection metrics—precision, recall, and mean Average Precision (mAP@0.50)—alongside training loss behavior, confusion matrices, and interpretability through Grad-CAM visualizations. Together, these evaluations provide a comprehensive assessment of model accuracy, robustness, and practical applicability in real-world agricultural environments..

YOLO Models Performance Analysis

The YOLOv8 and YOLOv9 models achieved competitive results across all evaluation metrics, reflecting their efficiency in real-time detection tasks. As shown in Table 1, YOLOv9 slightly outperformed YOLOv8, achieving an mAP@0.50 of 0.769 compared to 0.758 for YOLOv8. This improvement is attributed to YOLOv9's architectural refinements, which enhance feature representation while maintaining fast inference. Precision values for both models exceeded 0.68, while recall scores above 0.71 indicate that most rice panicles were successfully detected.

Table 1. YOLO Models Performance

Model Name	Precision (mP)	Recall (mR)	mAP@0.50
YOLOv8	0.68139	0.72792	0.75867
YOLOv9	0.69364	0.71375	0.76967

Loss Analysis

The training and validation loss curves (Fig. 3) confirm stable convergence for both YOLOv8 and YOLOv9. YOLOv9 demonstrated faster loss reduction and a smoother plateau, signifying better generalization with fewer overfitting tendencies. The gradual narrowing of the training-validation loss gap highlights the benefit of data augmentation strategies and transfer learning applied during model training..

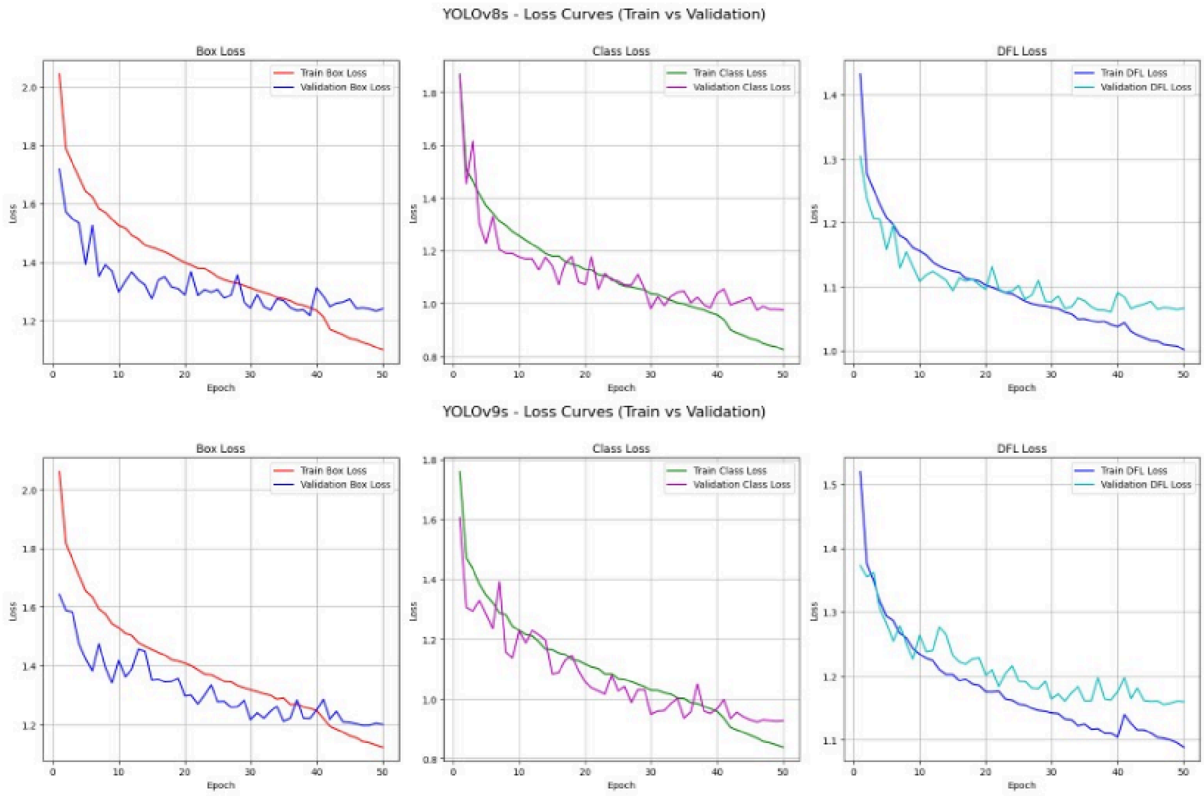

Fig. 3. YOLO Models Loss Curves

Confusion Matrix Analysis

The confusion matrices (Fig. 4-5) illustrate the models' classification accuracy across different rice panicle instances. Both YOLO models showed high diagonal dominance, indicating correct predictions in most cases. YOLOv9 displayed fewer false negatives than YOLOv8, which is consistent with its higher recall score. The misclassifications observed were mainly due to occluded panicles or images with extreme lighting variations, underscoring the complexity of field-level datasets..

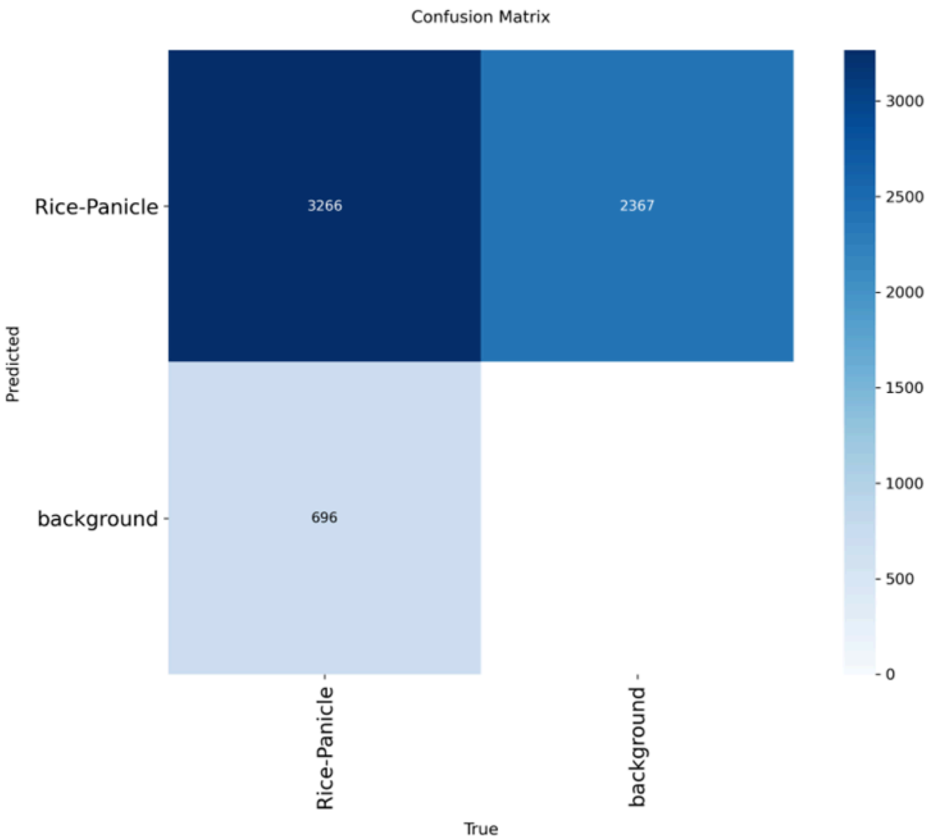

Fig. 4. Confusion Matrix of YOLOv8s Model

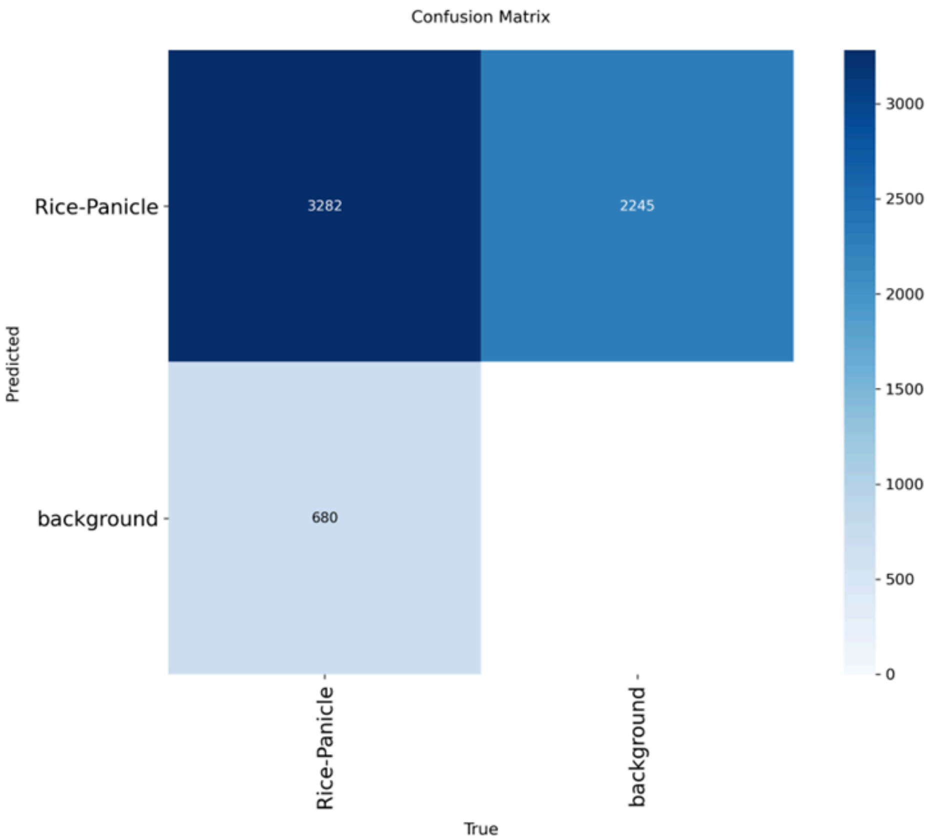

Fig. 5. Confusion Matrix of YOLOv8s Model

Model Interpretability

To enhance model transparency, Grad-CAM was applied to visualize the regions influencing the predictions (Fig. 6 and Fig. 7). Both YOLOv8 and YOLOv9 highlighted panicle clusters with strong attention focus, validating that the models concentrated on biologically meaningful regions. YOLOv9 demonstrated more precise localization, with heatmaps tightly covering panicle areas while minimizing background influence. This interpretability reassures that the detection results are not spurious, thereby increasing trust in real-world deployment..

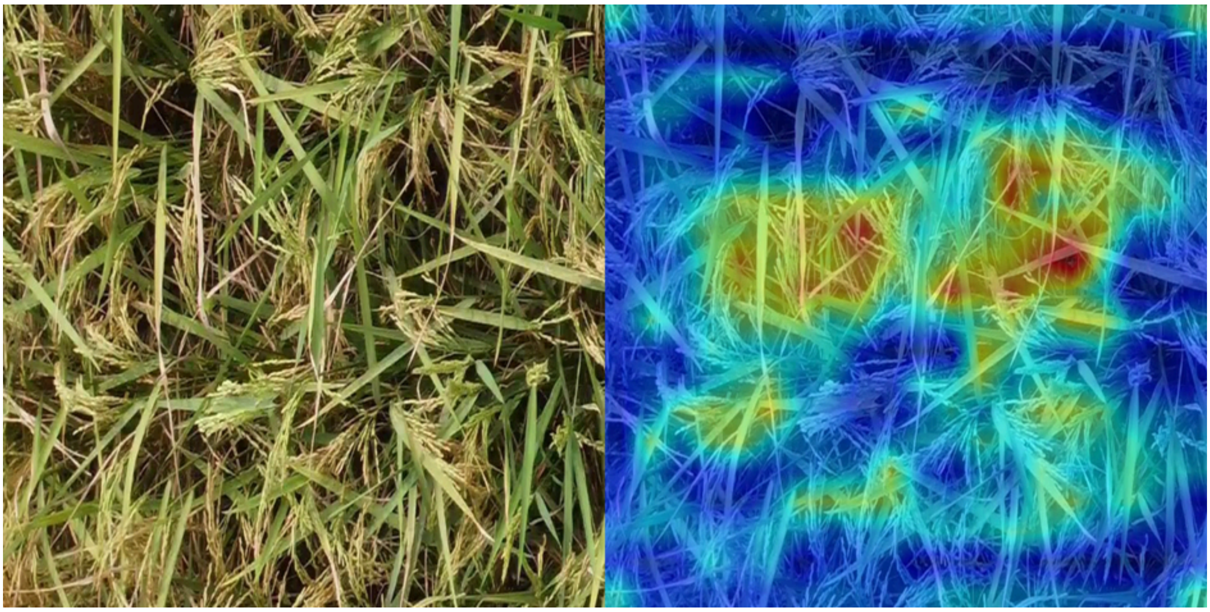

Fig. 6. YOLOv8s Model Interpretability

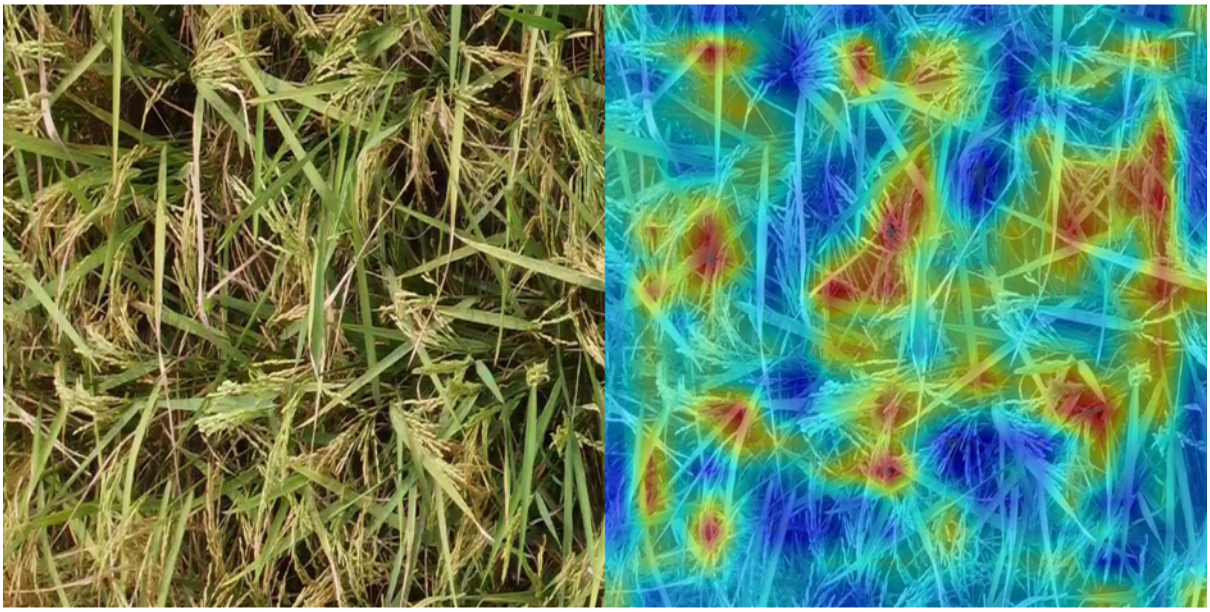

Fig. 7. YOLOv9s Model Interpretability

Self Supervised Models Performance

The DINO-YOLO series, which integrates self-supervised pre-training, provided additional insights into the adaptability of SSL techniques for agricultural datasets. As presented in Table 2, DINO-YOLOv9 achieved slightly higher recall (0.554) and mAP@0.50 (0.527) compared to DINO-YOLOv8. However, overall performance lagged behind the standard YOLO models, likely due to limited fine-tuning epochs and the complexity of transferring SSL features directly to dense detection tasks. Nevertheless, these results demonstrate the potential of SSL for reducing annotation requirements in future large-scale applications..

Table 2. DINO-YOLO Models Performance

Model Name	Precision (mP)	Recall (mR)	mAP@0.50
DINO-YOLOv8	0.5215	0.5321	0.5178
DINO-YOLOv9	0.5125	0.5540	0.5274

Principal Component Analysis

To further analyze feature representation, Principal Component Analysis (PCA) was applied to the embedding space of the DINO-YOLO models (Fig. 8). The clusters of panicle-related features were more compact in DINO-YOLOv9 compared to DINO-YOLOv8, reflecting its improved recall performance. However, the overlap between certain classes suggests the need for further domain-specific fine-tuning to maximize the benefits of SSL..

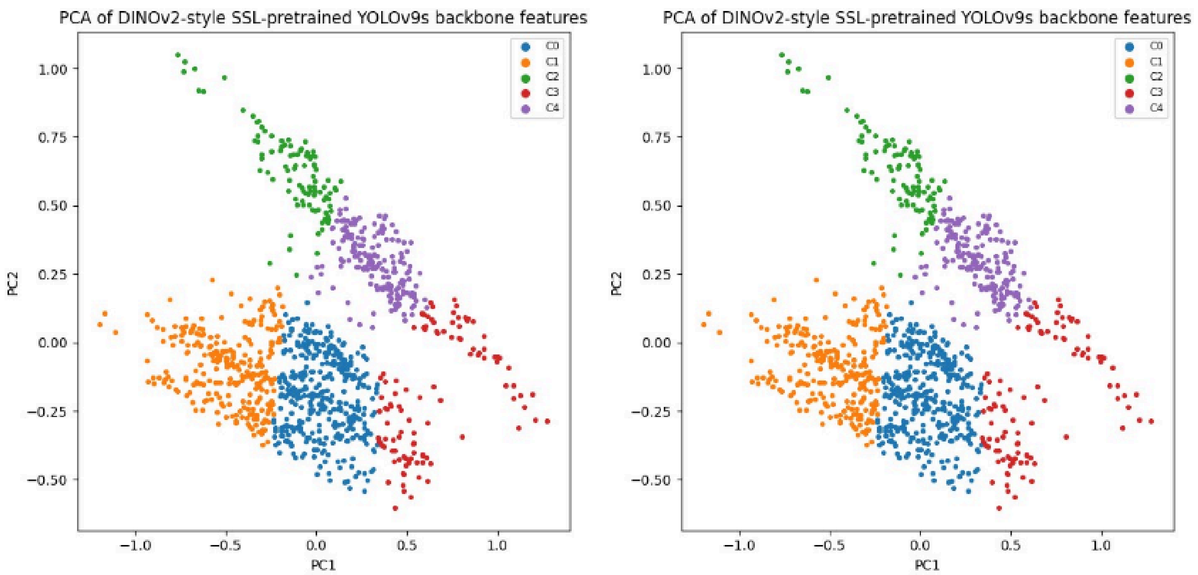

Fig. 8. DINO-YOLO Models PCA Analysis

Overall, the YOLO models—particularly YOLOv9—demonstrated superior accuracy and robustness for rice panicle detection in Bangladeshi field conditions. DINO-YOLO models, while underperforming in raw detection metrics, revealed the promise of SSL approaches in scenarios with limited annotated data. The Grad-CAM interpretability further confirmed that the models focused on biologically relevant regions, supporting their reliability for agricultural stakeholders. These results validate the feasibility of combining transfer learning, SSL, and XAI to advance precision agriculture

DISCUSSION

The experimental findings highlight the effectiveness of modern object detection frameworks for rice panicle detection in real-world Bangladeshi field conditions. In particular, YOLOv9 outperformed YOLOv8 across all evaluated metrics, reaffirming its improved feature extraction and localization capacity. The precision and recall scores of both models exceed those reported in earlier panicle detection studies indicating progress toward reliable, real-time detection in complex agricultural environments. The stable convergence observed in the loss analysis further suggests that the applied augmentation strategies and transfer learning significantly reduced overfitting risks, which is often a challenge in domain-specific datasets.

The confusion matrix analysis and Grad-CAM visualizations provided crucial interpretability, revealing that the models consistently focused on panicle-rich regions while disregarding irrelevant background noise. This transparency is particularly important in agriculture, where domain experts require assurance that AI-driven predictions are biologically meaningful. Compared with prior interpretability efforts in crop detection tasks, our results demonstrate stronger attention alignment with true panicle locations, making the framework more trustworthy for field deployment.

On the other hand, the performance of self-supervised DINO-YOLO models was modest relative to supervised YOLO variants. While DINO-YOLOv9 achieved slightly better recall and clustering of panicle features than DINO-YOLOv8, both models lagged behind YOLOv8 and YOLOv9 in precision and mAP. These findings highlight a key challenge of directly transferring self-supervised representations to dense object detection tasks. Nonetheless, the results validate the potential of SSL in agricultural domains where annotation resources are limited. Earlier studies have already emphasized SSL's utility in disease detection and plant trait analysis, and our results extend this evidence to panicle detection, albeit with room for improvement.

From a practical perspective, the integration of Weighted Boxes Fusion (WBF) with multiple models improved robustness by refining bounding box predictions, reducing false positives in occluded and cluttered field scenes. Such ensemble approaches have been shown to strengthen reliability in prior agricultural vision research, and our study confirms their value in panicle detection under variable lighting and background conditions.

Despite these promising results, several limitations must be acknowledged. First, the dataset, although carefully curated, was collected from a single geographic region (Gazipur, Bangladesh), which may restrict the generalizability of the trained models to other agroecological zones. Future work should extend the dataset to diverse field conditions, capturing seasonal variability, soil heterogeneity, and different rice cultivars. Second, while Grad-CAM provided interpretability at a coarse spatial level, finer-grained explanation methods such as SHAP or integrated gradients could yield deeper insights into feature attribution. Third, the modest performance of DINO-YOLO models highlights the need for longer fine-tuning schedules or hybrid SSL-supervised strategies to fully leverage self-supervised pre-training.

Overall, this study advances the state of rice panicle detection by combining high-performing YOLO architectures with explainable AI and exploring the potential of self-supervised learning. The results hold significant implications for precision agriculture in Bangladesh, where timely and accurate yield estimation directly impacts food security and resource allocation. By reducing reliance on manual panicle counting and introducing interpretable, automated detection pipelines, the proposed framework contributes toward scalable smart agriculture solutions..

CONCLUSION

This study proposed and evaluated a deep learning–based framework for automated rice panicle detection in Bangladeshi field conditions, integrating supervised YOLO models, self-supervised DINO-YOLO variants, and explainable AI techniques. The results demonstrated that YOLOv9 achieved the highest accuracy, with strong precision, recall, and mAP values, while also producing more stable training convergence compared to YOLOv8. Grad-CAM visualizations further confirmed that the YOLO models consistently attended to biologically meaningful panicle regions, enhancing interpretability and trustworthiness of predictions.

Although the self-supervised DINO-YOLO models underperformed relative to their fully supervised counterparts, they provided valuable insights into the adaptability of SSL in agricultural detection tasks. Their ability to cluster panicle features with limited annotation suggests strong potential for future applications where large-scale labeled datasets are unavailable. The use of Weighted Boxes Fusion (WBF) further improved robustness, reducing false detections in challenging field scenes.

The findings underscore the feasibility of deploying AI-driven detection systems for rice yield monitoring in Bangladesh, reducing dependence on manual surveys and supporting timely decision-making for farmers and policymakers. Future work will focus on expanding the dataset across diverse agroecological zones, refining SSL pre-training strategies, and exploring multimodal data sources (e.g., UAV, hyperspectral imagery) to further improve generalization. By combining accuracy, interpretability, and adaptability, the proposed framework contributes meaningfully to advancing precision agriculture and strengthening national food security initiatives..

REFERENCES

1. Hamed, M. A., El-Habib, M. F., Sababa, R. Z., Al-Hanjor, M. M., Abunasser, B. S., & Abu-Naser, S. S. (2024). Artificial intelligence in agriculture: Enhancing productivity and sustainability.

2. M. Tan, R. Pang, Q. V. Le, "EfficientDet: Scalable and efficient object detection," *CoRR*, abs/1911.09070, 2019. [Online]. Available: http://arxiv.org/abs/1911.09070.

3. X. Wang, W. Yang, Q. Lv, C. Huang, X. Liang, G. Chen, L. Xiong, L. Duan, "Field rice panicle detection and counting based on deep learning," *Frontiers in Plant Science*, vol. 13, 2022. doi: 10.3389/fpls.2022.966495. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2022.966495.

4. S. Xiang, S. Wang, M. Xu, W. Wang, W. Liu, "YOLO Pod: A fast and accurate multi-task model for dense soybean pod counting," *Plant Methods*, vol. 19, p. 4, 2023. doi: 10.1186/s13007-023-00985-4.

5. P. Sun, J. Cui, X. Hu, Q. Wang, "WDN: A one-stage detection network for wheat heads with high performance," *Information*, vol. 13, no. 3, p. 153, 2022. doi: 10.3390/info13030153. [Online]. Available: https://www.mdpi.com/2078-2489/13/3/153.

6. A. Koirala, K. B. Walsh, Z. Wang, N. Anderson, "Deep learning for mango (Mangifera indica) panicle stage classification," *Agronomy*, vol. 10, no. 1, p. 143, 2020. [Online]. Available: https://www.mdpi.com/2073-4395/10/1/143.

7. M. S. Hossain, M. Rahman, A. Rahman, M. M. Kabir, M. F. Mridha, J. Huang, J. Shin, "Automatic navigation and self-driving technology in agricultural machinery: A state-of-the-art systematic review," *IEEE Access*, 2025.

8. G. Yilma, M. Dagne, M. K. Ahmed, R. B. Bellam, "Attentive Self-supervised Contrastive Learning (ASCL) for plant disease classification," *Results in Engineering*, vol. 25, p. 103922, 2025.

9. A. Mamun, D. Ahmedt-Aristizabal, M. Zhang, M. I. Hossen, Z. Hayder, M. Awrangjeb, "Plant disease detection using self-supervised learning: A systematic review," *IEEE Access*, 2024.

10. Z. Wang, R. Wang, M. Wang, T. Lai, M. Zhang, "Self-supervised transformer-based pre-training method with General Plant Infection dataset," in *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Singapore, 2024, pp. 189-202,

Springer Nature Singapore.

11. N. Yang, K. Chang, J. Tang, L. Xu, Y. He, R. Huang, J. Yu, "Detection method of rice blast based on 4D light field refocusing depth information fusion," *Computers and Electronics in Agriculture*, vol. 205, p. 107614, 2023. doi: 10.1016/j.compag.2023.107614. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923000029.

12. L. Yu, J. Shi, C. Huang, L. Duan, D. Wu, D. Fu, C. Wu, L. Xiong, W. Yang, Q. Liu, "An integrated rice panicle phenotyping method based on X-ray and RGB scanning and deep learning," *The Crop Journal*, vol. 9, no. 1, pp. 42-56, 2021. doi: 10.1016/j.cj.2020.06.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214514120301094.

13. W. Hu, C. Zhang, Y. Jiang, C. Huang, Q. Liu, L. Xiong, W. Yang, F. Chen, "Nondestructive 3D image analysis pipeline to extract rice grain traits using X-ray computed tomography," *Plant Phenomics*, 2020. doi: 10.34133/2020/3414926. [Online]. Available: https://spj.science.org/doi/abs/10.34133/2020/3414926.

14. S. V. Desai, V. N. Balasubramanian, T. Fukatsu, S. Ninomiya, W. Guo, "Automatic estimation of heading date of paddy rice using deep learning," *Plant Methods*, vol. 15, no. 1, p. 57, 2019. doi: 10.1186/s13007-019-0457-1. [Online]. Available: https://doi.org/10.1186/s13007-019-0457-1.

15. R. Deng, Y. Jiang, M. Tao, X. Huang, K. Bangura, C. Liu, J. Lin, L. Qi, "Deep learning-based automatic detection of productive tillers in rice," *Computers and Electronics in Agriculture*, vol. 177, p. 105703, 2020. doi: 10.1016/j.compag.2020.105703. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169920310462.

16. M. Sun, S. Huang, Z. Lu, M. Wang, S. Zhang, K. Yang, B. Tang, W. Yang, C. Huang, "A novel method for intelligent analysis of rice yield traits based on LED transmission imaging and cloud computing," *Measurement*, vol. 217, p. 113017, 2023.

17. Y. Guo, S. Li, Z. Zhang, Y. Li, Z. Hu, D. Xin, Q. Chen, J. Wang, R. Zhu, "Automatic and accurate calculation of rice seed setting rate based on image segmentation and deep learning," *Frontiers in Plant Science*, vol. 12, p. 770916, 2021. doi: 10.3389/fpls.2021.770916. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2021.770916.

18. C. P. J. Singh, "Paddy grain maturity estimation using deep learning approach-Yolomaturity," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 9, 2023.

19. J. Zhou, X. Lu, R. Yang, Y. Wang, H. Chen, J. Shen, M. Chen, Z. Zhou, F. Liu, "Developing thermal infrared de-ghost and multi-level nested conglutinated segmentation algorithm for detection of rice seed setting rate," *Computers and Electronics in Agriculture*, vol. 207, p. 107725, 2023. doi: 10.1016/j.compag.2023.107725. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923001138.

20. M. Lyu, X. Lu, Y. Shen, Y. Tan, L. Wan, Q. Shu, Y. He, Y. He, H. Cen, "UAV time-series imagery with novel machine learning to estimate heading dates of rice accessions for breeding," *Agricultural and Forest Meteorology*, vol. 341, p. 109646, 2023.

21. A. Barreto, P. Lottes, F. R. I. Yamati, S. Baumgarten, N. A. Wolf, C. Stachniss, A.-K. Mahlein, S. Paulus, "Automatic UAV-based counting of seedlings in sugar-beet field and extension to maize and strawberry," *Computers and Electronics in Agriculture*, vol. 191, p. 106493, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169920310462.

22. G. Ziyue, C. Yang, W. Yang, G. Chen, Z. Jiang, B. Wang, J. Zhang, "Panicle ratio network: Streamlining rice panicle measurement by deep learning with ultra-high-definition aerial images in fields," *Journal of Experimental Botany*, 2022. doi: 10.1093/jxb/erac294. [Online]. Available: https://doi.org/10.1093/jxb/erac294.

23. R. Deng, L. Qi, W. Pan, Z. Wang, D. Fu, X. Yang, "Automatic estimation of rice grain number based on a convolutional neural network," *J. Opt. Soc. Am. A*, vol. 39, no. 6, pp. 1034-1044, 2022. doi: 10.1364/JOSAA.459580. [Online]. Available: https://opg.optica.org/josaa/abstract.cfm?URI=josaa-39-6-1034.

24. C. Zhou, H. Ye, J. Hu, X. Shi, S. Hua, J. Yue, Z. Xu, G. Yang, "Automated counting of rice panicle by applying deep learning model to images from unmanned aerial vehicle platform," *Sensors*, vol. 19, no. 14, p. 3106, 2019. doi: 10.3390/s19143106. [Online]. Available: https://www.mdpi.com/1424-8220/19/14/3106.

25. W. Guo, T. Fukatsu, S. Ninomiya, "Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images," *Plant Methods*, vol. 11, p. 7, 2015. doi: 10.1186/s13007-015-0047-9.

26. K. Liu, J. Wang, K. Zhang, M. Chen, H. Zhao, J. Liao, "A lightweight recognition method for rice growth period based on improved YOLOv5s," *Sensors*, vol. 23, no. 15, p. 6738, 2023. doi: 10.3390/s23156738. [Online]. Available: https://www.mdpi.com/1424-8220/23/15/6738.

27. G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael, T. Xie, J. Fang, "YOLOv5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation," Nov. 2022.

[Online]. Available: https://doi.org/10.5281/zenodo.7347926.

28. C.-Y. Wang, A. Bochkovskiy, H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv:2207.02696, 2022. [Online]. Available: https://arxiv.org/abs/2207.02696.

29. G. Jocher, A. Chaurasia, J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics.

30. R. A. Solovyev, W. Wang, "Weighted boxes fusion: Ensembling boxes for object detection models," CoRR, vol. abs/1910.13302, 2019. [Online]. Available: http://arxiv.org/abs/1910.13302.

31. J. Ci, E. J. van Henten, X. Wang, A. K. Burusa, G. Kootstra, "SSL-NBV: A self-supervised-learning-based next-best-view algorithm for efficient 3D plant reconstruction by a robot," *Computers and Electronics in Agriculture*, vol. 233, p. 110121, 2025. doi: 10.1016/j.compag.2025.110121.

32. Y. Cao, G. Sun, Y. Yuan, L. Chen, "Small-sample cucumber disease identification based on multimodal self-supervised learning," *Crop Protection*, vol. 188, p. 107006, 2025.

33. Y. Wang, Y. Yin, L. Li, T. Qu, Z. Guo, M. Peng, F. Li, "Classification of plant leaf disease recognition based on self-supervised learning," *Agronomy*, vol. 14, no. 3, p. 500, 2024.

34. A. Mamun, D. Ahmedt-Aristizabal, M. Zhang, M. I. Hossen, Z. Hayder, M. Awrangjeb, "Plant disease detection using self-supervised learning: A systematic review," *IEEE Access*, 2024.

35. Y. Wang, Y. Yin, L. Li, T. Qu, Z. Guo, M. Peng, F. Li, "Classification of plant leaf disease recognition based on self-supervised learning," *Agronomy*, vol. 14, no. 3, p. 500, 2024.

36. R. Zhao, Y. Zhu, Y. Li, "CLA: A self-supervised contrastive learning method for leaf disease identification with domain adaptation," *Computers and Electronics in Agriculture*, vol. 211, p. 107967, 2023. doi: 10.1016/j.compag.2023.107967.

37. S. Tan, H. Lu, J. Yu, M. Lan, X. Hu, H. Zheng, Y. Peng, Y. Wang, Z. Li, L. Qi, X. Ma, "In-field rice panicles detection and growth stages recognition based on riceRes2Net," *Computers and Electronics in Agriculture*, vol. 206, p. 107704, 2023. doi: 10.1016/j.compag.2023.107704. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169923000923.

38. R. Deng, M. Tao, X. Huang, K. Bangura, Q. Jiang, Y. Jiang, L. Qi, "Automated counting grains on the rice panicle based on deep learning method," *Sensors*, vol. 21, no. 1, p. 281, 2021. doi: 10.3390/s21010281. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/281.

39. L. Wei, Y. Luo, L. Xu, Q. Zhang, Q. Cai, M. Shen, "Deep convolutional neural network for rice density prescription map at ripening stage using unmanned aerial vehicle-based remotely sensed images," *Remote Sensing*, vol. 14, no. 1, p. 46, 2022. doi: 10.3390/rs14010046. [Online]. Available: https://www.mdpi.com/2072-4292/14/1/46.

40. C. Zhou, H. Ye, J. Hu, X. Shi, S. Hua, J. Yue, Z. Xu, G. Yang, "Automated counting of rice panicle by applying deep learning model to images from unmanned aerial vehicle platform," *Sensors*, vol. 19, no. 14, p. 3106, 2019. doi: 10.3390/s19143106. [Online]. Available: https://www.mdpi.com/1424-8220/19/14/3106.

41. M. R. A. Rashid, M. S. Hossain, M. Fahim, M. S. Islam, R. H. Prito, M. S. A. Sheikh, M. S. Ali, M. Hasan, M. Islam, et al., "Comprehensive dataset of annotated rice panicle image from Bangladesh," *Data in Brief*, vol. 51, p. 109772, 2023. doi: 10.1016/j.dib.2023.109772.

APPENDIX

Notebook and Links:

YOLOv8s: https://www.kaggle.com/code/raiyangani/cse521p-yolov8-xai

YOLOv9s: https://www.kaggle.com/code/raiyangani/cse521p-yolov9-xai

DINO-YOLOv8: https://www.kaggle.com/code/raiyangani/cse521p-dino-yolov8

DINO-YOLOv9: https://www.kaggle.com/code/raiyangani/cse521p-dino-yolov9

BYOL-YOLO: https://www.kaggle.com/code/raiyangani/cse521p-byol-yolo

DINO-RF-DETR: https://www.kaggle.com/code/raiyangani/cse521p-dinov2-rf-detr