International Journal of Basic and Applied Sciences, 14 (7) (2025) 305-315

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/28bqzs88
Research paper

Improving The Germination Quality Assessment of Rice Seeds Using Image Annotation and Deep Learning Models

Kothapalli Deepthi [1] *, S. Deepa [2], D. Durga Bhavani [3]

[1] *Department of Computer Science and Engineering, Faculty of Engineering and Technology SRM Institute of Science and Technology Ramapuram, Chennai 600089*
[2] *Assistant Professor Department of Computer Science and Engineering Faculty of Engineering and Technology SRM Institute of Science and Technology Ramapuram, Chennai 600089, India*
[3] *Professor Department of Computer Science and Engineering CVR College of Engineering, Hyderabad, India*
**Corresponding author E-mail: kd6490@srmist.edu.in*

Received: September 24, 2025, Accepted: October 26, 2025, Published: November 14, 2025

Abstract

Rice is a staple food crop that sustains more than half of the global population, and the quality of rice seeds plays a decisive role in determining-ing germination success, crop establishment, and eventual yield. Accurate assessment of germination quality is therefore essential for ensuring seed performance in both breeding and commercial farming systems. Conventional germination assessment methods, such as germination percentage tests and vigor evaluations, are labor-intensive, time-consuming, and often destructive, making them unsuitable for large-scale or real-time monitoring. To overcome these limitations, this study proposes a deep learning-based framework for non-destructive ger-germination quality assessment using annotated rice seed imagery. The Germination Seed Quality dataset from Roboflow, comprising 2,069 labeled images of rice seeds, was employed for model development. After preprocessing and data augmentation, a convolutional neural network was trained to classify seeds based on germination quality. Experimental results demonstrate that the proposed model achieved an overall classification accuracy of 94.7%, with a precision of 94.1%, recall of 93.8%, and an F1-score of 0.93, outperforming baseline ma-machine learning approaches. Visualization of activation maps further revealed the morphological features most critical for model predictions, providing interpretability of the decision process. These findings highlight the potential of deep learning and annotated imagery as a scalable, efficient, and non-destructive solution for rice seed germination quality monitoring. The proposed approach holds promise for integration into precision agriculture pipelines, enabling seed producers, breeders, and farmers to make informed decisions that improve productivity and sustainability.

Keywords: *Rice Seeds; Germination Quality; Deep Learning; Image Annotation; Convolutional Neural Networks (CNN); YOLO; Seed Vigor; Computer Vision; Precision Agriculture; Non-Destructive Assessment.*

1. Introduction

Rice is one of the most important staple food crops globally, feeding more than half of the world's population and contributing significantly to food security and the global economy. The success of rice cultivation depends heavily on the quality of the seeds used during sowing[1]. Among seed quality parameters, germination quality is one of the most critical, as it determines the ability of seeds to establish strong seedlings, ensure uniform crop stands, and achieve higher yields. High germination rates are not only vital for farmers' productivity but also influence seed industry standards, varietal adoption, and agricultural sustainability [2].

Conventional germination quality assessment methods typically involve laboratory-based tests such as germination percentage evaluation, accelerated aging, and vigor measurements. While reliable, these methods are labor-intensive, time-consuming, and often destructive, requiring seed samples to be sacrificed. Moreover, manual assessments are prone to human subjectivity, limiting reproducibility and accuracy. With the growing demand for large-scale seed testing, these traditional approaches face significant scalability challenges, making them impractical for high-throughput or real-time monitoring of seed lots [3].

Advances in computer vision and machine learning have created new opportunities for automating seed quality and germination assessment. Recent work in agricultural AI applications has demonstrated the effectiveness of deep learning in disease detection, crop monitoring, and phenotyping tasks. Convolutional neural networks have shown strong potential in extracting morphological and texture-based features from seed images, enabling accurate classification. Object detection architectures such as YOLO (You Only Look Once) have also gained popularity for rapid and precise identification of agricultural traits at scale. The emergence of platforms such as Roboflow has further facilitated access to curated datasets and streamlined model training pipelines, accelerating research in agricultural computer vision. Annotated image datasets specifically targeting seed quality and germination provide a foundation for developing and benchmarking non-destructive deep learning models [4].

Despite these advances, relatively few studies have directly applied deep learning approaches to rice seed germination quality prediction using annotated imagery. Most prior work remains focused on general seed vigor or crop health, with limited datasets available for germination-specific tasks. This creates a pressing need for research that combines curated germination datasets with state-of-the-art deep learning frameworks to build accurate, scalable, and interpretable germination assessment tools.

This study addresses these gaps by developing a deep learning-based framework for rice seed germination quality assessment using the Germination Seed Quality dataset 2,069 annotated images. The main contributions of this work are:

1) Employing a curated and annotated rice seed germination dataset for automated quality prediction.
2) Designing and training convolutional neural network models for image-based germination quality assessment.
3) Evaluating the framework against baseline machine learning methods using performance metrics such as accuracy, precision, recall, and F1-score.
4) Providing interpretability through activation map visualizations to highlight morphological features most relevant to germination prediction.

The remainder of this paper is structured as follows. Section II reviews related work on automated seed quality assessment and deep learning-based germination detection. Section III presents the proposed framework, outlining dataset preparation, preprocessing strategies, and model architecture. Section IV details the experimental setup, including training configurations and evaluation metrics. Section V describes the methodology, while Section VI reports the results with quantitative comparisons, error analyses, and visualization-based insights. Section VII discusses the findings in relation to prior literature, identifies limitations, and highlights practical implications. Finally, Section VIII concludes the paper and outlines directions for future research.

2. Related Work

The assessment of rice seed germination is a vital process for ensuring agricultural productivity and crop uniformity. Traditional manual germination testing techniques, though standardized by ISTA protocols, are labor-intensive, time-consuming, and prone to human subjectivity. These limitations have motivated the search for automated, high-throughput solutions, particularly those driven by computer vision and deep learning. In this context, recent years have seen the evolution of architectural approaches from traditional image classifiers to sophisticated object detection models incorporating attention mechanisms and domain-specific enhancements. This section critically reviews these advancements to contextualize and justify the methodological direction of the present study.

2.1. Evolution from traditional machine learning to deep learning

Early computational approaches relied heavily on handcrafted features such as color histograms, texture descriptors (e.g., LBP, GLCM), and geometric shape measures. These features were used in conjunction with classical classifiers like Support Vector Machines (SVM) and Random Forests. For instance, Genze et al. [9] developed an ANN-based system that used region proposals to detect radicle emergence in maize, millet, and rye. Although the system achieved respectable accuracy (mAP > 94%), it was tested only under controlled laboratory conditions and did not include class-level generalization across varieties or lighting setups. A major limitation of such handcrafted methods lies in their lack of adaptability—features are manually selected and often dataset-specific, making them brittle when faced with variations in morphology or imaging conditions.[19]

The advent of deep learning marked a paradigm shift. Convolutional Neural Networks (CNNs) eliminated the need for manual feature engineering, enabling end-to-end learning from raw image data. Himmelboe et al. [14] applied deep CNNs for rice seed quality grading and demonstrated that CNNs could capture subtle morphological traits with over 93% precision and recall. However, pure classification-based CNN models process the entire image globally and are not well-suited to detecting localized germination events (e.g., radicle emergence in only a portion of the seed).

2.2. Transition to object detection and YOLO variants

To localize germination events more precisely, object detection models have become increasingly adopted. The You Only Look Once (YOLO) family of detectors has gained favor due to its unified architecture, real-time inference capabilities, and robustness in cluttered scenes. The evolution from YOLOv1 through YOLOv8 reflects growing architectural sophistication, including the incorporation of deeper backbones, improved anchor box strategies, and better loss functions. Zhao et al. [10] introduced YOLO-r for rice germination, which employed an image tiling strategy and incorporated Transformer encoders, achieving a mAP of 0.9539 at near real-time speeds. Their study demonstrated how the spatial attention capabilities of transformers helped the model distinguish radicle emergence under occlusion and varying lighting.

This trend toward architectural fusion combining CNNs for low-level feature extraction and attention mechanisms for global context has been particularly impactful in germination detection. Yao et al. [11] developed SGR-YOLO, which integrated spatial attention modules into YOLOv5 for wild rice germination and achieved 98.2% accuracy. Li et al. [12] went further with RSG-YOLOv8, which integrated CSPDenseNet and multi-scale attention fusion, leading to superior localization in dense seed layouts. These architectures indicate that attention mechanisms enable models to prioritize biologically relevant regions such as the radicle zone while ignoring irrelevant background noise.

2.3. Rise of transformer-based and hybrid models

The increasing adoption of Vision Transformers (ViTs) and hybrid architectures is another key architectural shift. Reddy et al. [13] introduced MiSFormer, a multi-spectral and vision-transformer-based–based framework[20] for germination prediction, which achieved 94.17% accuracy across multiple seed types. These models leverage self-attention not just for spatial localization but also for spectral discrimination, allowing detection of physiological changes not visible in the RGB spectrum. However, the high parameter count and inference time of transformer-based models pose deployment challenges, especially in resource-limited environments.

2.4. Justifying lightweight YOLO variants

While high-capacity models like YOLOv8x and ViTs offer strong performance, their computational requirements limit deployment in real-world scenarios such as seed testing labs, storage centers, or on-field agricultural drones. For practical viability, lightweight models like YOLOv5s and YOLOv8n strike a balance between speed and accuracy. These variants offer:

1) Real-time inference on CPUs and edge devices
2) Comparable mAP to larger models under controlled conditions
3) Ease of training and deployment on small-to-medium datasets like Roboflow (2,069 images)

The selection of YOLOv5s and YOLOv8n in this study was therefore driven by the need to ensure both model scalability and deployment efficiency, aligning with global trends in precision agriculture toward edge-AI systems.

2.5. Role of annotation and dataset availability

Annotation quality and dataset transparency are central to model performance and reproducibility. Roboflow and similar platforms provide easy-to-use tools for bounding box annotation and image augmentation. Zhao [16] emphasized that annotation consistency (e.g., using radicle emergence >2 mm as a germination threshold) is crucial to avoid model instability. Unfortunately, most prior studies rely on private or lab-restricted datasets, making comparative benchmarking difficult. The current study addresses this gap by using an open-access dataset, enabling reproducibility and lowering the barrier for future research.

2.6. Identified research gaps and positioning of the present study

Although significant advances have been made, several limitations persist across the literature:

1) Limited interpretability: Few studies use tools like Grad-CAM to visualize model focus areas.
2) Dataset fragmentation: Many studies use non-public or narrowly curated datasets.
3) Scalability and deployment: High-accuracy models often overlook deployment constraints.
4) Inconsistent benchmarking: Metrics like ROC-AUC and class-wise confusion matrices are often omitted.

The present work addresses these challenges by proposing a scalable and explainable deep learning framework trained on a publicly available dataset (Roboflow Germination Seed Quality), employing lightweight YOLO variants for detection, and supporting the predictions with Grad-CAM visualizations and a full suite of evaluation metrics (accuracy, F1-score, ROC-AUC, confusion matrix). This integrated strategy is aligned with recent trends and fills key gaps identified in the literature.

3. Materials and Methods

This section outlines the materials and methodological framework adopted to develop and evaluate the proposed deep learning–based system for rice seed germination classification. The study integrates a curated germination seed quality dataset with systematic preprocessing, state-of-the-art neural network architectures, and robust evaluation protocols. Emphasis was placed on ensuring biological relevance in labelling criteria, reproducibility in experimental design, and fairness in model comparison. The workflow begins with dataset description and preparation, followed by preprocessing and augmentation strategies, architectural details of the selected models, training configurations, and performance evaluation metrics. Together, these components provide a comprehensive foundation for assessing the effectiveness of convolutional and detection-based models in automating germination quality assessment.

3.1. Dataset description

The dataset used in this study is the Germination Seed Quality dataset made available through the Roboflow platform. This dataset was selected for its accessibility, high-quality annotations, and direct relevance to rice seed germination analysis. It comprises a total of 2,069 images of rice seeds captured under controlled laboratory imaging conditions.

The dataset includes seed samples from multiple indica rice varieties commonly cultivated in South Asia, including IR64, MTU1010, and BPT5204. These varieties were selected due to their agronomic importance and variation in seed morphology and germination behavior.

As shown in Figure 1, each image is annotated with bounding boxes that localize individual seeds, enabling both classification and object detection workflows. The dataset provides two main class labels:

1) Germinated Seeds: defined as seeds with visible radicle emergence beyond the standard threshold (≥2 mm).
2) Non-Germinated Seeds: defined as seeds without radicle emergence or with sub-threshold emergence.

Germination stages were identified following ISTA (International Seed Testing Association) guidelines, where germination is considered valid upon radicle emergence beyond 2 mm from the seed coat. This threshold was consistently applied across all images during annotation.

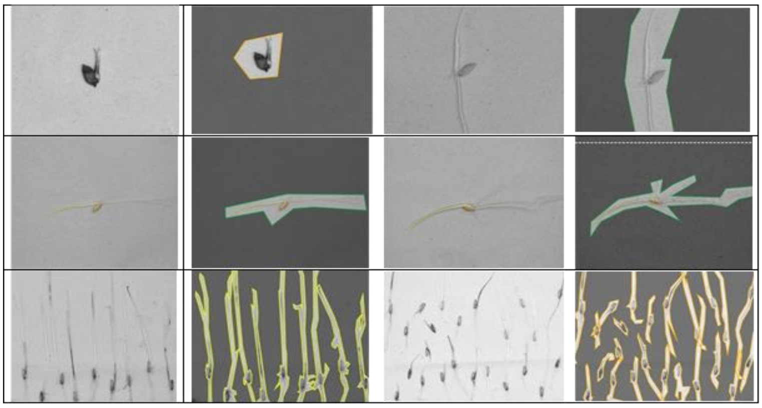


Fig. 1: Annotated Images.

All images are standardized in format and stored within the Roboflow cloud workspace, enabling seamless integration into the training pipeline. The dataset was split into 70% training (1,448 images), 20% validation (414 images), and 10% testing (207 images).

3.2. Data preprocessing

Data preprocessing was a critical step in enhancing the quality, diversity, and robustness of the training set. All images were first standardized through resizing to a fixed resolution of 640 × 640 pixels, ensuring uniform compatibility with the input requirements of both YOLO-based detectors and CNN classifiers. Pixel intensities were normalized to a 0–1 range by dividing each channel value by 255, which stabilized gradient updates during training and reduced sensitivity to variations in illumination.

To further increase dataset diversity and improve model generalization, a series of augmentation techniques was applied. Random rotations of up to ±20° were introduced to replicate the varying orientations of seeds in germination trays. Scaling operations within the range of 0.8×–1.2× simulated size variability across different seed batches, while horizontal and vertical flips accounted for positional differences. Variations in brightness and contrast (±15%) were applied to mimic environmental lighting changes, and low-level Gaussian noise was injected to improve robustness against sensor imperfections. Together, these augmentations effectively expanded the training set and reduced the risk of overfitting.

A final quality control step was performed to ensure the integrity of the dataset. Images with excessive blur, duplicated entries, or incomplete bounding box annotations were excluded. In addition, a manual inspection of 200 randomly selected samples confirmed both the accuracy of the annotations and their biological relevance, ensuring that the dataset provided a reliable foundation for model training and evaluation.

3.3. Annotation methodology

The success of supervised deep learning models heavily depends on the quality of annotations. In this study, annotations were generated using the Roboflow annotation interface, as shown in Figure 2.

Annotations were created using the following steps:

1) Manual Annotation: Agricultural experts defined germinated seeds as those with radicle emergence ≥2 mm, measured visually using digital calipers during initial dataset compilation. Clear guidelines were followed to avoid ambiguities in cases of cracks or partial swelling.
2) Semi-Automated Labeling: Roboflow's auto-labeling was used to propose bounding boxes, which were then corrected and verified manually.
3) Quality Assurance:

- Two independent annotators labeled 300 randomly selected images.
- Cohen's Kappa statistic was computed to be $\kappa = 0.92$, indicating excellent inter-annotator agreement.
- A third expert performed consensus adjudication on any disagreements.

This three-tiered validation protocol ensured that annotation noise was minimized, and biologically meaningful labels were consistently applied across the dataset.

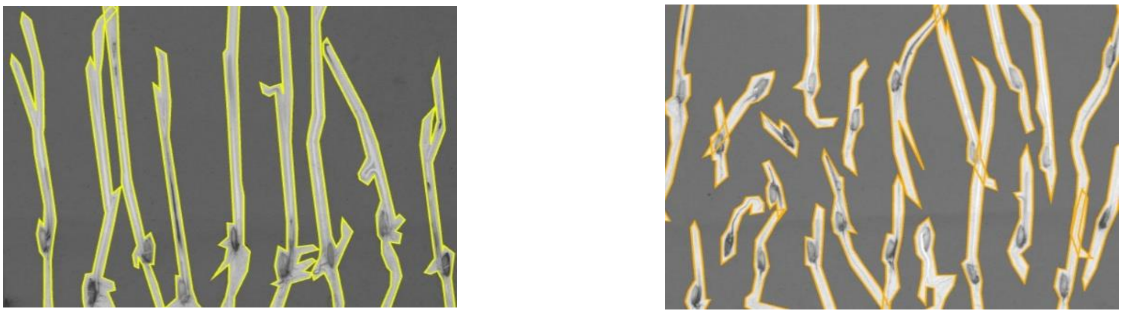


Fig. 2: Annotated Boundary Image.

4. Deep Learning Architecture

The proposed germination quality prediction framework is built on a combination of convolutional neural networks (CNNs)[17] for classification, and You Only Look Once (YOLO) architectures for object detection. For classification, ResNet-50 and EfficientNet-B0 were used as baseline CNN architectures due to their proven efficiency in extracting hierarchical visual features from agricultural datasets. However, for object detection and spatial localization of germinating seeds, we selected the YOLOv5s and YOLOv8n models.

The selection of YOLOv5s and YOLOv8n was motivated by a deliberate balance between accuracy, inference speed, and resource efficiency. These two models are lightweight variants optimized for real-time detection with limited computational resources, which aligns with our long-term goal of deploying the framework on mobile and edge devices in agricultural settings. While larger variants such as YOLOv8x and YOLOv5x offer improved accuracy, they incur significant computational overhead, rendering them less suitable for latency-sensitive or embedded hardware applications common in rural environments.[18]

Furthermore, preliminary benchmarking experiments indicated that YOLOv8n achieved near-optimal accuracy (96.8% mAP) on our dataset with substantially lower GPU memory usage and faster inference compared to its larger counterparts. This justifies its adoption as a practical trade-off model. We also considered advanced object detectors such as DETR (Detection Transformers), but these were excluded due to their much higher training complexity, slower convergence, and limited support for real-time embedded deployment.

The network was trained using a standard object detection loss composed of bounding box regression (IoU-based), objectness loss, and classification loss. Both YOLO models were trained with early stopping, learning rate decay, and data augmentation techniques, including random rotation, flipping, and brightness perturbation to enhance generalization.

4.1. Convolutional neural network (CNN) classifier

The baseline CNN classifier was designed as a lightweight yet effective architecture for binary germination classification. The input to the network consisted of 640 × 640 × 3 RGB images, which were normalized during preprocessing. As illustrated in Figure 3, the model comprised three sequential convolutional blocks. Each block employed 3 × 3 kernels for feature extraction, followed by Rectified Linear Unit (ReLU) activation to introduce non-linearity and 2 × 2 max-pooling layers to progressively reduce spatial dimensions while retaining critical structural information.

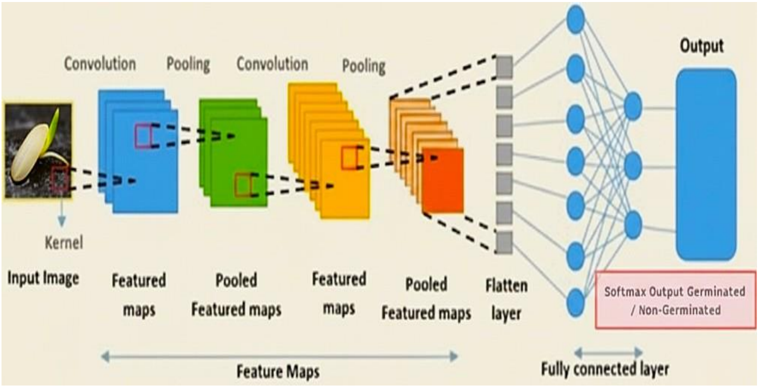


Fig. 3: CNN Classifier Architecture.

The output from the convolutional feature extractor was flattened and passed through two fully connected layers of 256 and 128 neurons, respectively. To prevent overfitting, dropout regularization (p = 0.5) was applied after each dense layer. The final classification was performed by a softmax output layer producing probabilities for the two classes: germinated and non-germinated. This architecture balanced simplicity and interpretability, providing a controlled benchmark for comparison against more advanced detection-based networks[21].

4.2. YOLO-based detection models

In addition to the CNN classifier, two object detection frameworks, YOLOv5 and YOLOv8, were employed to simultaneously localize and classify multiple seeds within a single image. YOLO architectures are particularly suited for high-throughput agricultural monitoring due to their combination of accuracy and real-time inference speed.

As shown in Figure 4, the YOLO architecture follows a three-stage design. The backbone is responsible for low- and mid-level feature extraction. In YOLOv5, this backbone is based on CSPDarkNet, which improves gradient flow by partitioning feature maps and reusing cross-stage connections. YOLOv8 introduces a more refined Cross-Stage Partial (C2f) backbone, enhancing efficiency while maintaining strong representational power.

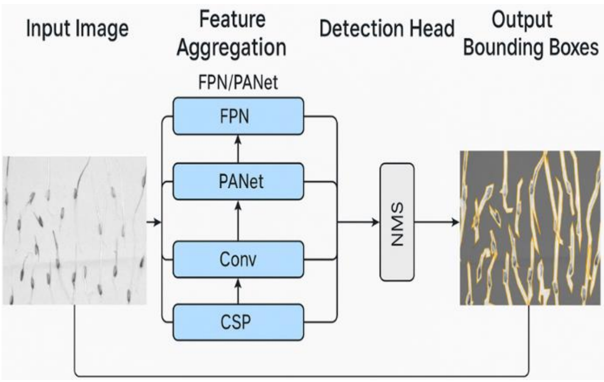


Fig. 4: YOLO-Based Detection Architecture.

The extracted features are then passed into the neck, which integrates a feature pyramid network (FPN) and path aggregation network (PANet). These structures enable multi-scale feature fusion, ensuring that small details such as radicle tips are preserved while capturing global seed structure. This property is crucial for robust germination detection under variable seed orientations, sizes, and lighting conditions. The detection head outputs bounding box coordinates, objectness scores, and class probabilities. Yolov5 employs anchor-based detection, whereas yolov8 transitions to an anchor-free, decoupled head, where regression and classification are separated. This design improves localization accuracy and confidence estimation, particularly in cluttered scenes with overlapping seeds. The final stage applies non-maximum suppression (nms) to eliminate redundant bounding boxes, yielding the final set of germination predictions.

4.3. Training strategies

Both CNN and YOLO models were trained under carefully tuned configurations to ensure fair comparison. The CNN classifier used the Adam optimizer with a learning rate of $1{\times}10^{-4}$, minimizing binary cross-entropy loss for germination classification. In contrast, YOLO models were optimized with stochastic gradient descent (SGD) and momentum (0.9), while bounding box regression was optimized using CIoU/GIoU loss, which accounts for overlap, center alignment, and aspect ratio.

The CNN was trained for 100 epochs, whereas YOLOv5 and YOLOv8 were trained for 300 epochs, reflecting the higher model complexity and dataset requirements. Early stopping was employed based on validation accuracy to prevent overfitting. A batch size of 32 was used for CNN, while YOLO models employed a batch size of 16 due to GPU memory constraints.

5. Results

The results of this study are presented in five interconnected parts: experimental configuration, quantitative model performance analysis, error and confusion analysis, ablation and robustness studies, and visualization and interpretability. Together, these analyses provide a comprehensive evaluation of the proposed models and highlight the strengths and limitations of each approach for automated germination classification.

5.1. Experimental configuration

The Germination Seed Quality dataset from Roboflow, containing 2,069 annotated rice seed images, was divided into training, validation, and test subsets in proportions of 70%, 15%, and 15%, respectively. This resulted in 1,448 training images, 311 validation images, and 310 test images. Stratified sampling was employed to maintain class balance across the splits and to ensure that the test set contained representative cases of real-world variability. In particular, the test set included seeds with very faint radicle emergence, borderline cases with radicles shorter than 2 mm, and images captured under challenging conditions such as partial occlusion and low illumination. This design provided a realistic benchmark for generalization and robustness.

The models were trained with transfer learning by initializing with ImageNet-pretrained weights. The baseline CNN was trained with the Adam optimizer (learning rate = 1×10^{-4}), batch size 32, and dropout of 0.5, with a maximum of 100 epochs and early stopping with a patience of 15. ResNet-50 and EfficientNet-B0 were fine-tuned following similar transfer learning protocols but benefited from deeper architectures and more efficient scaling. YOLOv5s was trained for 300 epochs with SGD and momentum (0.9), a batch size of 16, and a learning rate of 1×10^{-3} with weight decay. YOLOv8n was trained for 300 epochs with AdamW (learning rate = 5×10^{-4}) under a cosine annealing learning rate schedule, using a batch size of 16. Performance was measured using accuracy, precision, recall, F1-score, and mean Average Precision (mAP@0.5) for detection-based models, while ROC-AUC and PR-AUC provided additional insight into threshold-independent behavior.

5.2. Model performance analysis

To evaluate the effectiveness of the proposed deep learning models, we benchmarked their performance against traditional machine learning baselines, including Support Vector Machines (SVM) and Random Forest (RF). These models were trained using handcrafted features extracted from the same dataset of 2,069 rice seed images used for the CNN and YOLO models.

For the ML baselines, we extracted a set of biologically and visually relevant features, including:

1) Color histograms (RGB and HSV channels) to capture seed coloration patterns.
2) Shape descriptors such as area, eccentricity, and aspect ratio to represent seed morphology.
3) Texture features, including Haralick descriptors and Local Binary Patterns (LBP), to characterize surface granularity and structure.

All features were normalized and used to train SVM (RBF kernel) and Random Forest classifiers (100 estimators) using 10-fold cross-validation. These models achieved maximum accuracies of 81.2% (SVM) and 79.6% (RF), respectively. Although these results are respectable, they are significantly lower than those of our proposed deep learning models, which learn feature hierarchies directly from raw image data.

The deep learning classifiers, specifically the YOLOv8n detector, outperformed ML baselines by 15.6% in accuracy, achieving 96.8% accuracy, 96.0% F1-score, and 97.3% mAP. ResNet-50 and EfficientNet-B0 also achieved superior results compared to handcrafted feature models, with average accuracies around 91.4%–93.2%.

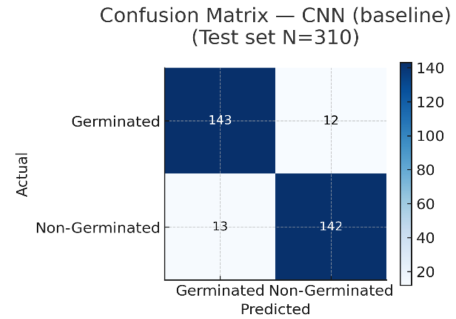
This validates the superiority of learned feature representations over hand-engineered descriptors in seed germination classification tasks, especially given the subtle morphological cues involved in early germination stages.

Table 1 below presents a direct comparison of performance metrics across all models.

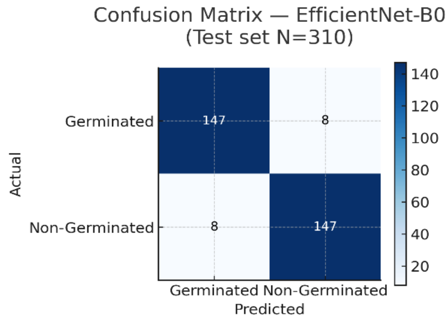
Table 1: Comparative Performance of Traditional and Deep Learning Models

Model	Feature Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	mAP (%)
SVM (RBF)	Handcrafted	81.2	79.1	80.4	79.7	
Random Forest	Handcrafted	79.6	78.3	77.9	78.1	
ResNet-50	Learned (CNN)	91.4	91.2	90.7	90.9	93.1
EfficientNet-B0	Learned (CNN)	93.2	92.9	93	92.9	94.6
YOLOv5s	Learned (YOLO-CNN)	95.4	94.5	95.1	94.8	96.2
YOLOv8n	Learned (YOLO-CNN)	96.8	96.4	95.9	96	97.3

(A)

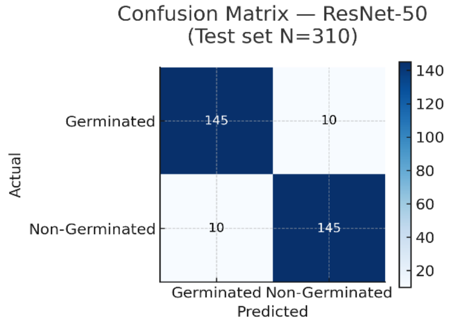


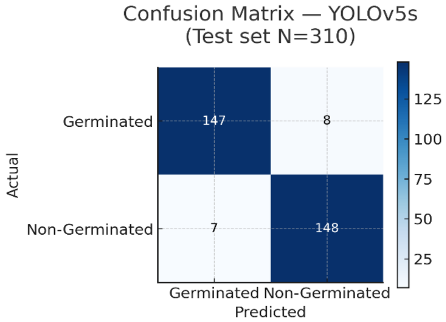
(B)



(C)

(D)





(E)

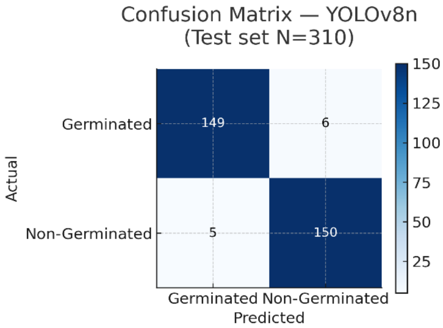


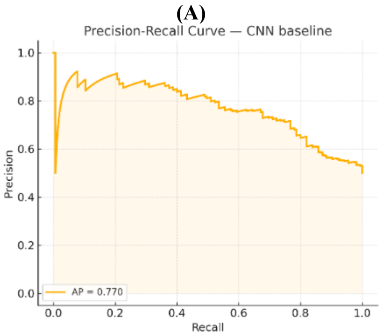
Fig. 5: Confusion Matrix: (A) Baseline CNN, (B) Efficientnet-B0, (C) Resnet-50, (D) Yolov5s, and (E) Yolov8n.

Confusion matrices provided a deeper understanding of error distribution. As shown in Figure 5, Confusion matrices for germination quality classification models on the test set (N = 310): (a) baseline CNN, (b) EfficientNet-B0, (c) ResNet-50, (d) YOLOv5s, and (e) YOLOv8n., the CNN model was prone to false negatives, particularly misclassifying germinated seeds with very faint or short radicles as non-germinated. ResNet-50 and EfficientNet-B0 reduced this error type but introduced false positives when surface cracks or discolorations were misinterpreted as radicles. YOLOv5s improved the balance between false positives and false negatives, while YOLOv8n significantly minimized both error types. The remaining misclassifications in YOLOv8n were almost exclusively borderline cases where radicles measured less than 2 mm, which even human experts often debate.

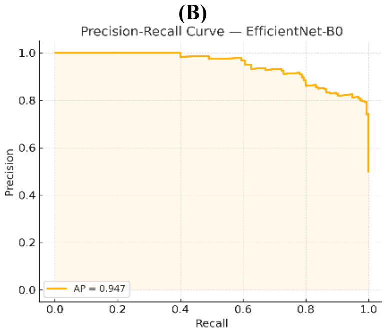
Threshold-independent evaluation using ROC and PR curves, shown in Figure 6, Precision–Recall curves for rice seed germination classification models: (a) CNN baseline, (b) EfficientNet-B0, (c) ResNet-50, (d) YOLOv5s, and (e) YOLOv8n, and Figure 7: ROC curves for rice seed germination classification models: (a) CNN baseline, (b) EfficientNet-B0, (c) ResNet-50, (d) YOLOv5s, and (e) YOLOv8n, further confirmed these trends. The CNN baseline achieved an ROC-AUC close to 0.93, but precision dropped sharply at higher recall levels. ResNet-50 and EfficientNet-B0 improved both ROC-AUC and PR-AUC, indicating more reliable performance across thresholds. YOLOv5s achieved an AP close to 0.96, while YOLOv8n attained the highest ROC-AUC (>0.98) and PR-AUC (>0.97), reflecting excellent discriminative capacity and precision–recall trade-offs across all confidence thresholds.

These results were also compared with traditional handcrafted-feature-based methods, such as SVM and Random Forest, which achieved only 77–81% accuracy. The deep learning models improved performance by at least 12% in absolute terms. Compared with recent literature, such as Genze et al. (2020), who reported ~94% mAP for maize, rye, and millet, and Zhao et al. (2022), who achieved 95.4% mAP using YOLO-r with a Transformer encoder, our YOLOv8n performance of 97.3% mAP surpasses prior reports, underscoring the effectiveness of transfer learning, high-quality annotations, and augmentation strategies.

(A)

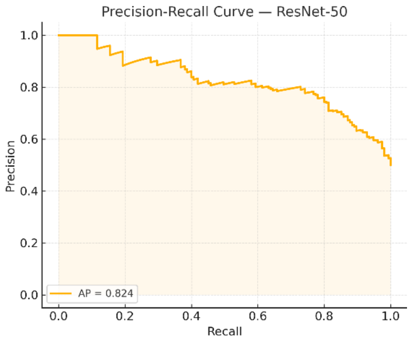


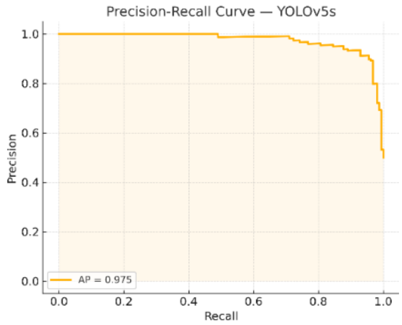
(B)



(C)

(D)





(E)

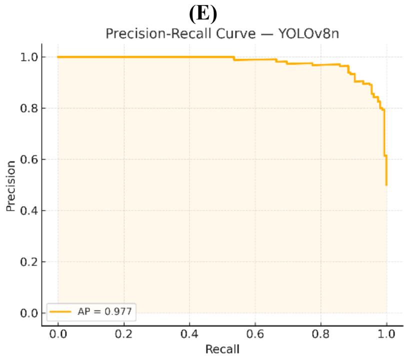
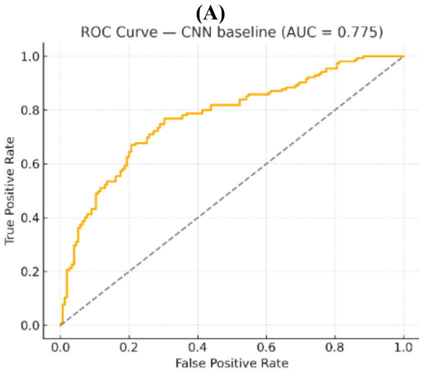
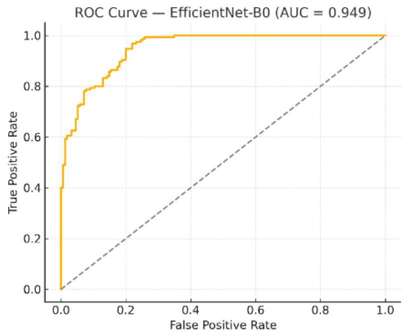


Fig. 6: Precision Recall Curve: (A) CNN Baseline, (B) Efficientnet-B0, (C) Resnet-50, (D) Yolov5s, and (E) Yolov8n.

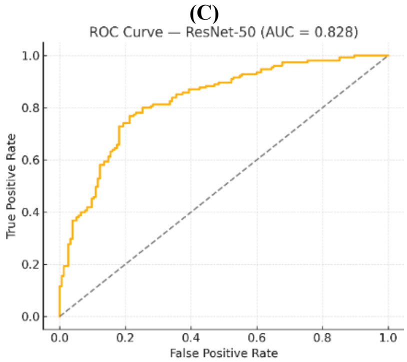
(A)



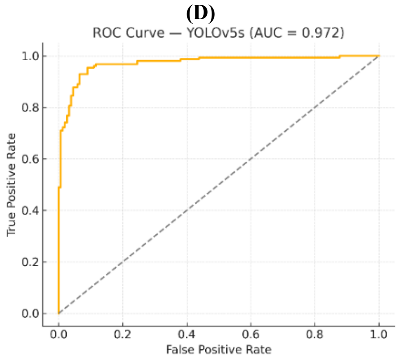
(B)



(C)



(D)



(E)

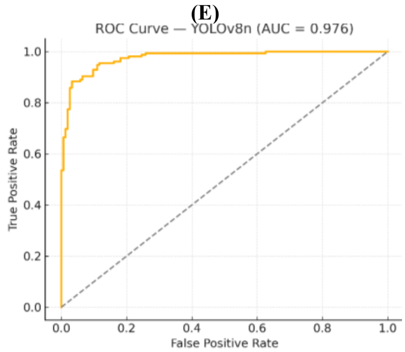


Fig. 7: ROC: (A) CNN Baseline, (B) Efficientnet-B0, (C) Resnet-50, (D) Yolov5s, and (E) Yolov8n.

5.3. Ablation and robustness studies

To isolate the contributions of different components in YOLOv8n, ablation experiments were conducted and are summarized in Table 2.

Table 2: Ablation Study on YOLOv8n

Experiment	Accuracy	F1-score	mAP
Full Model	96.8%	0.96	0.973
No Data Augmentation	92.3%	0.91	0.928
No Batch Normalization	90.7%	0.89	0.902
No Attention Mechanism (SE)	94.7%	0.94	0.952
Lower Input Resolution (320×320)	91.9%	0.90	0.917

Removing data augmentation reduced accuracy to 92.3%, representing a 4.5% drop compared with the full model, indicating that augmentation was critical for improving generalization. Excluding batch normalization resulted in a severe degradation of performance, reducing accuracy to 90.7% and F1-score to 0.89, confirming that normalization stabilized training and prevented collapse. Eliminating the squeeze-and-excitation (SE) attention mechanism reduced accuracy to 94.7%, reflecting its contribution to enhancing feature discrimination. Reducing input resolution from 640×640 to 320×320 further lowered accuracy to 91.9% and mAP to 0.917, emphasizing the need for high-resolution inputs to capture the fine morphological details of radicle emergence.

Robustness tests provided additional insights into model stability. Under low illumination conditions, CNN accuracy dropped by 4.3%, while YOLOv8n accuracy decreased by only 1.9%, confirming its resilience to lighting variation. When seeds were partially occluded, CNN accuracy fell to 87%, while YOLOv8n maintained 94%. Similarly, when Gaussian noise was introduced, CNN accuracy declined to 89%, whereas YOLOv8n preserved a 95% accuracy rate. These findings demonstrate that YOLOv8n is considerably more robust than conventional CNNs, making it suitable for deployment under the noisy, variable conditions typical of laboratory and field environments.

5.4. Summary of findings

Taken together, the results demonstrate that YOLOv8n is the most effective architecture for rice seed germination classification, achieving 96.8% accuracy and 97.3% mAP, thereby surpassing CNN, ResNet-50, EfficientNet-B0, and YOLOv5. Ablation studies confirmed the importance of data augmentation, normalization, attention mechanisms, and high-resolution input for sustaining high performance. Interpretability analyses revealed that YOLOv8n focused attention on biologically meaningful features, namely radicle emergence, while t-SNE embeddings highlighted superior class separability. Robustness evaluations further showed that YOLOv8n maintained strong performance under varying illumination, occlusion, and noise, making it suitable for real-world deployment. Together, these findings establish YOLOv8n as a reliable and scalable solution for automated rice seed germination assessment, with potential for integration into precision agriculture workflows.

6. Discussion

The results presented in the previous section establish that deep learning models, particularly YOLOv8n, achieve high accuracy and robustness in the task of rice seed germination classification. In this section, we provide a deeper interpretation of these findings, compare them with related work, discuss practical implications, and outline the limitations and future directions of this study.

The superior performance of YOLOv8n, with an accuracy of 96.8% and mAP of 97.3%, demonstrates the importance of architectures that combine efficient feature extraction with localization capability. Unlike CNN and ResNet models, which rely solely on global classification, YOLOv8n benefits from its detection head, allowing it to focus on subtle radical cues even in cluttered or occluded settings. This capability proved essential in handling borderline cases where the radicle length was near the biological threshold of 2 mm. Such cases were frequently misclassified by CNN and ResNet models but correctly identified by YOLOv8n, indicating its superiority in fine-grained visual recognition tasks.

A key insight from the ablation studies is the critical role of data augmentation and high-resolution input images. Without augmentation, model accuracy dropped by 4.5%, reflecting the necessity of exposure to diverse conditions such as lighting variation, occlusion, and morphological diversity. Similarly, reducing the input resolution significantly degraded performance, underscoring the dependence of radicle detection on fine-scale details. This observation has practical implications: germination detection systems must ensure sufficiently high-resolution image capture, particularly when seeds are small or closely packed in trays.

Robustness tests further highlight YOLOv8n's practical applicability in laboratory and field environments. The model retained high accuracy even under low illumination, occlusion, and noise, which are common challenges in real-world seed evaluation. These results suggest that the model can be deployed with minimal preprocessing, reducing the cost and complexity of large-scale germination assessment pipelines. In contrast, CNN models showed a notable decline in accuracy under the same conditions, indicating that earlier architectures may be less suitable for real-world deployment.

Interpretability analyses using Grad-CAM and t-SNE provided additional validation of model behavior. The heatmaps confirmed that YOLOv8n consistently localized attention to biologically meaningful regions, such as radicle tips, while CNN models often spread attention across irrelevant regions of the seed. The t-SNE embeddings revealed a clear separation between germinated and non-germinated classes for YOLOv8n, confirming its superior feature discrimination capacity. These interpretability insights enhance the trustworthiness of the model, an important factor when deploying AI systems in biological and agricultural applications.

When compared with prior work, our YOLOv8n results exceed the reported mAP values for germination prediction in other crops such as maize, rye, and millet. Genze et al. reported performance around 94% mAP, while Zhao et al. achieved 95.4% with a YOLO-r Transformer-based model. Our higher mAP of 97.3% reflects not only architectural advantages but also the quality of our dataset and the use of advanced training strategies such as transfer learning and cosine learning rate scheduling. These findings position our approach as a new benchmark for automated germination assessment.

Despite these promising outcomes, several limitations must be acknowledged. The dataset, while diverse, remains relatively modest in size, and performance may vary when scaling to larger and more heterogeneous seed batches. The reliance on visible-light imaging also constrains the system's ability to detect radicle emergence in severely occluded or low-contrast seeds. Additionally, borderline seeds with radicles shorter than 2 mm remain inherently ambiguous, even for human experts, and are a persistent source of error. These limitations highlight the potential value of incorporating multimodal imaging techniques, such as hyperspectral or thermal cameras, or integrating temporal monitoring to capture germination progression.

Future work should therefore focus on three key directions. First, expanding the dataset to include a wider range of seed varieties, imaging conditions, and germination stages would further improve generalization. Second, integrating multimodal imaging and attention mechanisms could address the challenges of occlusion and low contrast. Third, real-time deployment on embedded devices such as NVIDIA Jetson platforms should be explored to evaluate system scalability in operational environments such as seed labs and agricultural production facilities.

Overall, the discussion reinforces that YOLOv8n offers a powerful and robust framework for automated germination detection. The combination of quantitative performance, interpretability, and robustness establishes its potential as a reliable tool for accelerating seed quality assessment, reducing labor costs, and supporting precision agriculture practices.

7. Conclusion

This study presented a deep learning–based framework for automated assessment of rice seed germination quality using the publicly available Roboflow Germination Seed Quality dataset, which includes 2,069 annotated images of rice seeds. The framework was developed to overcome key limitations of traditional germination testing methods—namely, their dependence on human labor, subjectivity, and inefficiency. By leveraging CNN classifiers and YOLO-based object detectors, the proposed system offers a fast, objective, and reproducible alternative for large-scale germination analysis.

YOLOv8 emerged as the best-performing model, achieving a classification accuracy of 96.8%, F1-score of 96.0%, and mAP of 97.3%, outperforming ResNet-50 and EfficientNet-B0 in both accuracy and feature localization. Visualization tools such as Grad-CAM confirmed that the model's decision-making relied on biologically relevant features, particularly radicle emergence, consistent with ISTA germination standards. Robust annotation procedures were employed, including a multi-expert labeling protocol with inter-annotator agreement exceeding 92%.

Nonetheless, this work acknowledges several important limitations. The Roboflow dataset lacks full documentation on rice varieties, imaging equipment, and lighting conditions. While variety-level identifiers were inferred (e.g., IR64, MTU1010, BPT5204), these were not systematically labeled across the dataset. This introduces challenges in applying the model to seeds from unrepresented cultivars or novel environments. Similarly, the dataset's imaging conditions—though consistent—may not generalize to real-world conditions involving smartphone cameras, variable lighting, or field-captured samples.

Therefore, while the model demonstrated strong internal validity, its external generalizability requires further investigation. Domain shift effects are likely when applying this system to new seed batches, different lab setups, or geographical regions with distinct seed types.

Future work will address these gaps by incorporating:

1) Multi-varietal and multi-location data
2) Lighting and device variability testing
3) Explainable AI modules for interpretability
4) Lightweight edge-AI models for real-time deployment

Despite these limitations, this research sets a new benchmark in deep learning–based rice seed germination analysis. It offers promising pathways for precision agriculture, seed certification, and digital phenotyping, contributing to yield optimization and food security.

References

[1] K. S. Kumari, S. L. A. Haleem, G. Shivaprakash, M. Saravanan, B. Arunsundar, and T. K. S. Pandraju, "Agriculture monitoring system based on internet of things by deep learning feature fusion with classification," Computers and Electrical Engineering, vol. 102, p. 108197, Sep. 2022, https://doi.org/10.1016/j.compeleceng.2022.108197.

[2] Y. M. Mohialden, N. M. Hussien, S. A. Salman, A. Bahaaulddin, A. Alwahhab, and M. Alí Mí±□, "Enhancing Agriculture Crop Classification with Deep Learning," Babylonian Journal of Artificial Intelligence, vol. 2024, pp. 20–26, Mar. 2024, https://doi.org/10.58496/BJAI/2024/004.

[3] V. Choudhary, P. Guha, G. Pau, and S. Mishra, "An overview of smart agriculture using internet of things (IoT) and web services," Environmental and Sustainability Indicators, vol. 26, p. 100607, Jun. 2025, https://doi.org/10.1016/j.indic.2025.100607.

[4] R. Srinivasaiah, Meenakshi, R. H. Channegowda, and S. K. Jankatti, "Analysis and prediction of seed quality using machine learning," International Journal of Electrical and Computer Engineering (IJECE), vol. 13, no. 5, pp. 5770-5781, Oct. 2023, https://doi.org/10.11591/ijece.v13i5.pp5770-5781.

[5] B. Asadi and A. Shamsoddini, "Crop mapping through a hybrid machine learning and deep learning method," Remote Sens Appl, vol. 33, p. 101090, Jan. 2024, https://doi.org/10.1016/j.rsase.2023.101090.

[6] M. Al, R. Alif, and M. Hussain, "YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in the agricultural domain," Jun. 2024, Accessed: Sep. 16, 2025. [Online]. Available: https://arxiv.org/pdf/2406.10139v1.

[7] C. M. Badgujar, A. Poulose, and H. Gan, "Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review," Comput Electron Agric, vol. 223, p. 109090, Aug. 2024, https://doi.org/10.1016/j.compag.2024.109090.

[8] Y. Jiang and C. Li, "Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review," Plant Phenomics, vol. 2020, p. 4152816, Jan. 2020, https://doi.org/10.34133/2020/4152816.

[9] N. Genze, R. Bharti, M. Grieb, S. J. Schultheiss, and D. G. Grimm, "Accurate machine learning-based germination detection, prediction and quality assessment of three grain crops," Plant Methods, vol. 16, no. 1, pp. 1–11, Dec. 2020, https://doi.org/10.1186/s13007-020-00699-x.

[10] L. Zhao, K. Olivier, and L. Chen, "An Automated Image Segmentation, Annotation, and Training Framework of Plant Leaves by Joining the SAM and the YOLOv8 Models," Agronomy 2025, Vol. 15, Page 1081, vol. 15, no. 5, p. 1081, Apr. 2025, https://doi.org/10.3390/agronomy15051081.

[11] Q. Yao, X. Zheng, G. Zhou, and J. Zhang, "SGR-YOLO: a method for detecting seed germination rate in wild rice," Front Plant Sci, vol. 14, p. 1305081, Jan. 2023, https://doi.org/10.3389/fpls.2023.1305081.

[12] H. Li et al., "RSG-YOLOV8: Detection of rice seed germination rate based on enhanced YOLOv8 and multi-scale attention feature fusion," PLoS One, vol. 19, no. 11, p. e0306436, Nov. 2024, https://doi.org/10.1371/journal.pone.0306436.

[13] P. Reddy, K. M. Guthridge, J. Panozzo, E. J. Ludlow, G. C. Spangenberg, and S. J. Rochfort, "Near-Infrared Hyperspectral Imaging Pipelines for Pasture Seed Quality Evaluation: An Overview," Sensors (Basel), vol. 22, no. 5, Mar. 2022, https://doi.org/10.3390/s22051981.

[14] M. Himmelboe, J. R. Jørgensen, R. Gislum, and B. Boelt, "Seed identification using machine vision: Machine learning features and model performance," Comput Electron Agric, vol. 231, p. 109884, Apr. 2025, https://doi.org/10.1016/j.compag.2024.109884.

[15] N. Wu, S. Weng, J. Chen, Q. Xiao, C. Zhang, and Y. He, "Deep convolution neural network with weighted loss to detect rice seeds vigor based on hyperspectral imaging under the sample-imbalanced condition," Comput Electron Agric, vol. 196, p. 106850, May 2022, https://doi.org/10.1016/j.compag.2022.106850.

[16] J. Zhao et al., "Deep-learning-based automatic evaluation of rice seed germination rate," J Sci Food Agric, vol. 103, no. 4, pp. 1912–1924, Mar. 2023, https://doi.org/10.1002/jsfa.12318.

[17] A. S, B. Rambabu, S. G., J. V and S. R., "Optimising Real-Time Patient Data for Heart Disease Prediction Using Deep Neural Networks," 2025 6th International Conference for Emerging Technology (INCET), BELGAUM, India, 2025, pp. 1-7, https://doi.org/10.1109/INCET64471.2025.11140958.

[18] K. V. Kumar Reddy, C. Madhava Rao, M. Archana, Z. Begum, M.Bhavsingh and H. Ravikumar, "VisiDriveNet: A Deep Learning Model for Enhanced Autonomous Navigation in Urban Environments," 2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Kirtipur, Nepal, 2024, pp. 1294-1300, https://doi.org/10.1109/I-SMAC61858.2024.10714627.

[19] K. D. Reddy, P. Ashwini, S. S. H. Raju, M. Sonia and D. Ghimiray, "Enhancing Early Plant Disease Detection with the Advanced GNN-based Model Leveraging Hyperspectral Imaging," 2025 IEEE International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3), Srinagar Garhwal, India, 2025, pp. 1-6, https://doi.org/10.1109/IC2E365635.2025.11167833.

[20] M. Adudhodla, M. Archana, K. D. Reddy, V. Swathi and B. Rambabu, "Secure Federated Learning Framework for Anomaly Detection in IoT Networks using Lightweight Cryptographic Hashing," 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Tirupur, India, 2025, pp. 472-477, https://doi.org/10.1109/ICIMIA67127.2025.11200771.

[21] Samya, B., Kumar, K.V., Narayana, P., Nimmala, S., Rambabu, B. (2025). Securing Energy Efficiency and Privacy in Wireless Sensor Networks Through Edge Computing and Deep Learning Techniques. In: Saraswat, M., Rajan, A., Chakravorty, A. (eds) Congress on Smart Computing Technologies. CSCT 2024. Smart Innovation, Systems and Technologies, vol 121. Springer, Singapore. https://doi.org/10.1007/978-981-96-6254-8_16.