YOLO-E3CA: An Ensemble YOLOv8 Framework with Coordinate Attention for Automated Detection of Karanda (*Carissa carandas*) Leaf Diseases

Amit Kumar Ghosh, Md Majidul Kabir, Shahriar Marjan*, Deepu Bhowmik, Rejowan Arifin Nayeem, Md Assaduzzaman

Department of Computer Science and Engineering, Daffodil International University, Dhaka-1216, Bangladesh
amitkumar89155@gmail.com, {kabir15-4812, marjan15-5126*, bhowmik15-5063, nayeem15-5108, assaduzzaman.cse}@diu.edu.bd

Abstract. Karanda (*Carissa carandas*), a significant tropical fruit crop of South Asia, faces remarkable yield and quality losses due to foliar fungal and bacterial diseases. Traditional methods of disease detection are slow, subjective, and inaccurate, resulting in late interventions and significant agricultural losses. This study presents an early curated and annotated Karanda Leaf Disease Datasets to advance research in automated disease detection. We present YOLO-E3CA (YOLOv8 Ensemble of 3 with Coordinate Attention), a novel network structure that consists of three different YOLOv8 version models with various hyperparameters, combined by a Coordinate Attention mechanism for enhancing spatial and channel feature representation. The model employs an advanced augmentation pipeline simulating diverse weather conditions and a custom training strategy to ensure robust generalization under noisy and fuzzy environments. With 98.5% mean average precision (mAP@50), YOLO-E3CA provides an accurate early diagnosis with a real-time web interface for practical accessibility. This study is the first to propose detecting Karanda leaf disease using ensemble learning and an attention mechanism in order to manage crops sustainably. In the future, we expect to diversify the deployment of our dataset and develop a mobile application for on-site inspection, providing affordable solutions for precision agriculture in extreme environmental conditions.

Keywords: Deep learning, Precision agriculture, Object detection, YOLOv8n, Coordinate attention.

1 Introduction

Karanda (*Carissa carandas*) is a tropical fruit and an economically important crop, being cultivated across the SAARC countries. Karanda Fruit is a rich natural source of vitamins, antioxidants and essential nutrients that cater various significant health-promoting and nutritional benefits [1]. Karanda provides livelihoods for millions of smallholder farmers and therefore it plays an important role in the regional and national economy. However, the crop is highly susceptible to a range of fungal and bacterial

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_59

diseases which cause major damage to leaves and fruits resulting in extremely low yields and quality [2]. The majority of traditional manual methods are subjective, time-consuming, and suffer from human errors especially in low-resources agricultural settings.

Recently, remarkable performance has been achieved in the automatic detection of plant disease using Artificial Intelligence (AI) and Deep Learning (DL)[3]. Adoption of deep learning models has the potential to perform well in extracting complex visual features from leaf images which allows an accurate and scalable disease classification compared to conventional methods. With the growing application of AI in agricultural research, several articles have studied fruit disease identification. However, no research has focused on leaf diseases of *Carissa carandas* (Karanda). To fill this gap, we have developed a Karanda leaf disease dataset with a YOLO based detection model for automated classification and segmentation of these diseases [4]. The system ensures high accuracy, robustness and real-time detection capacity, providing efficient agricultural monitoring and management.

In this work, a modified YOLOv8-based detection model is proposed for automated detection of Karanda leaf diseases. The proposed model was trained using our curated Karanda disease dataset that we prepared for this study. The modified YOLOv8 configuration significantly surpassed its standard counterpart, reaching mAP@50 of 98.5% and 96.3%, respectively, with an absolute difference of around 2.2%. This enhancement is owing to advanced data augmentation, coordinate attention incorporation, and a tailored training technique promoting robustness and generalization in the wild. In general, the proposed model can provide a more accurate and reliable estimate of disease detection performance, which will be of great benefit for precision agriculture and early disease management by farmers.

In addition, the developed YOLO-based detection model is completely embedded to a web platform for fast diagnosis of Karanda (*Carissa carandas*) leaf diseases without involving much processing time. The platform allows any user such as farmers, researchers and agricultural experts to load leaf images and get immediate real-time detection and classification. This deployment ensures availability, easy scaling and ease of use especially in resource poor settings. Accordingly, the research lays groundwork for an operational system that enables automated daily monitoring of Karanda crop health in real time, which will aid precision farming and sustainable Agro management.

2 Literature Review

In recent years, big strides have been made in deep learning and computer vision for autonomous plant disease detection, a long-standing problem that affects agricultural productivity and food security. By coupling object detection frameworks (in particular YOLO variants) with attention mechanisms, ensemble learning and more advanced augmentation strategies, the performance of disease identification on various crops has been drastically improved. In this work, we review the literature of 2025-state-of-art to investigate architectural enhancements in YOLOv8 based model architectures and ensemble methods and attention mechanism that led to improvements on real-world

agricultural datasets. The surveyed papers present the trends of multi-scale feature operations, model interpretability and efficiency optimization for mobile and edge-coding deployment in precision agriculture tasks.

Table 1. Summary of Related Works on Plant Disease Detection.

References	Objectives	Methods	Results	Limitations
Miao et al. [5]	Enhanced YOLOv8 for accurate plant disease detection across 13 species and 27 classes	SerpensGate-YOLOv8 with Dynamic Snake Convolution (DySnakeConv), SPPELAN module, and Super Token Attention (STA) mechanism	Precision: 71.9%; mAP@0.5: 64.9% (3.3% improvement over baseline YOLOv8); effective detection of elongated and twisted disease structures	Dataset imbalance affects generalization for rare diseases; misclassification between morphologically similar diseases; environmental factors (lighting, occlusion) impact performance
Ashurov et al. [6]	Automated plant disease identification using modified depthwise CNN with enhanced feature extraction	Depthwise CNN integrated with Squeeze-and-Excitation (SE) blocks and improved residual skip connections	Accuracy: 98%; F1 score: 98.2%; effective computational efficiency with real-time applicability	Data imbalance requires advanced augmentation; limited model interpretability; needs expansion to diverse plant species and diseases
Aboneh et al. [7]	Early crop disease detection for improved food security using ensemble learning and computer vision	Ensemble model combining Inceptionv3, ResNet50, and VGG19 pre-trained models on 23,000+ crop images	Training accuracy: 96.14%; validation accuracy: 96.85%; improved generalizability and robustness	Model explainability challenges; scalability concerns; drift issues in deployment environments
Bo Gan et al. [8]	Lightweight and efficient YOLOv8 detection head for resource-constrained agricultural deployment	Enhanced YOLOv8 with lightweight detection head optimized for mobile and edge devices	Reduced manual intervention; automatic feature learning for plant diseases; improved inference speed	Computational constraints on low-end devices; limited performance on small disease regions; complex background interference
Mallick et al. [9]	Deep learning-based crop disease detection using transfer learning and object detection	CNN with transfer learning and YOLO for real-time disease localization and classification	High classification accuracy with minimal training samples; accurate disease region annotation	Small-scale and unbalanced datasets affect performance; requires extensive augmentation strategies

References	Objectives	Methods	Results	Limitations
Korade et al. [10]	Tomato leaf disease detection with YOLOv8 and leaf extraction techniques	YOLOv8 with specialized leaf extraction preprocessing for tomato disease identification	Improved detection of stunted growth and bushy appearance symptoms; effective on mature plants	Limited to tomato species; specialized preprocessing reduces generalizability to other crops

3 Methodology

In this study, we develop YOLO-E3CA, a novel deep learning framework for automated detection of Karanda (*Carissa carandas*) leaf diseases, addressing the critical need for rapid and accurate diagnosis in precision agriculture. Our methodology, as illustrated in Fig 1, leverages a curated Karanda Leaf Disease Dataset, advanced data augmentation, and an ensemble of YOLOv8 models enhanced with Coordinate Attention to achieve robust performance across diverse environmental conditions. This section outlines the systematic processes of data acquisition, preprocessing, annotation, augmentation, and model architecture design that underpin our approach.

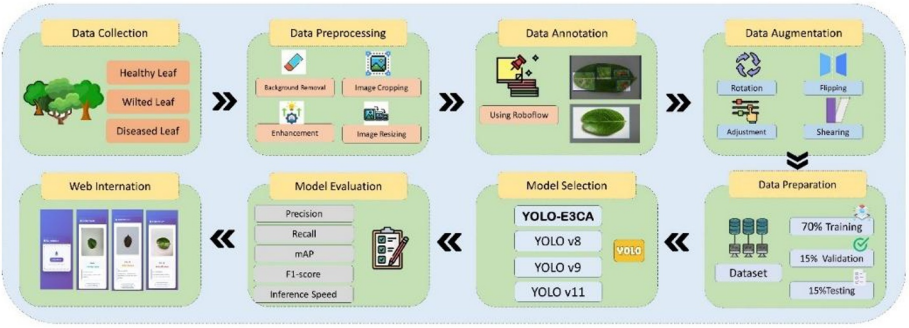

Fig. 1. Workflow of the proposed methodology.

3.1 Data Acquisition

We developed an image dataset of Karanda (*Carissa carandas*) leaves with various health conditions for design of a computerized system for detection of disease in karanda leaves. A collection of 2106 high quality raw images from January 2025 to June 2025 were collected for different regions of Bangladesh applying several devices. There are various classes features of the dataset: Healthy Leaf, Wilted Leaf, Diseased Leaf (Fig. 2). All the photos were taken under natural daylight conditions and real-world variations and image definition were kept. Images were taken using Samsung Galaxy S23 and iPhone 11 smartphones to maximize quality and resolution. This

database is useful in training and testing machine learning models for precise diagnosis and classification of Karanda leaf diseases.

Fig. 2. Karanda leaf dataset with 3 class: (a) Diseased Leaf, (b) Wilted Leaf, (c) Healthy Leaf.

3.2 Data Preprocessing

The dataset was preprocessed systematically for deep learning applications to standardize the images prior to model fitting. The image preprocessing involved the removal of uninteresting background areas, image contrast enhancement, and resizing images while maintaining their initial aspect ratio. These operations were necessary in order to decrease noise, enhancing the visual clarity of crucial features and setting input size to a consistent value across all cases. This normalization improved the quality of the dataset by which better and more stable models could be trained and evaluated.

3.3 Data annotation

After preprocessing, images were annotated to identify regions of disease via Roboflow. We used both automatic facilitation and manually validation to produce correct labels. We pre-processed this annotated dataset for training and testing the YOLO-based detection model.

3.4 Data Augmentation

The augmentation pipeline implemented in the code is a stochastic, multi-modal image transformation scheme designed to increase robustness to real-world imaging variation. It composes several mutually exclusive groups (via OneOf) and independent perturbations: a blur group (MotionBlur, MedianBlur, Blur; blur_limit=3) and an enhancement group (CLAHE, Sharpen, Emboss, RandomBrightnessContrast), each applied with moderate probability (p=0.3), followed by color-space perturbation through HueSaturationValue (hue ±20, saturation ±30, value ±20, p=0.3). Environmental effects are simulated with RandomFog (fog coefficient 0.1–0.3, p=0.1) and RandomSunFlare targeted to the upper image region (flare_roi=(0,0,1,0.5), p=0.1), while CoarseDropout (up to 8 rectangular holes, max 32×32 pixels, p=0.2) introduces localized occlusions and missing-pixel artifacts. By combining photometric, geometric-like occlusion, and weather-style corruptions in a probabilistic manner, the pipeline aims to expose the

detector to diverse appearance and noise conditions during training, thereby promoting better generalization to degraded, occluded, or variably illuminated scenes.

3.5 Model Architecture

The foundational backbone employs YOLOv8n's CSPDarknet architecture, which extracts hierarchical feature representations through convolutional layers with SiLU activation. The activation function is mathematically defined as:

$$SiLU(x) = x \cdot \sigma(x) = \frac{x}{1+e^{-x}} \tag{1}$$

where $\sigma(x)$ represents the sigmoid function. The backbone processes input images $I \in \mathbb{R}^{H\times W\times 3}$ through multiple C2f (Cross-Stage Partial with Fast connection) modules, generating multi-scale feature maps $F_i \in \mathbb{R}^{H_i\times W_i\times C_i}$ at different spatial resolutions.

The feature extraction follows:

$$F_i = \mathrm{C2f}_i\big(\mathrm{Conv}(F_{i-1})\big) \tag{2}$$

where Conv denotes the convolutional operation with batch normalization. The Spatial Pyramid Pooling Fast (SPPF) module enhances receptive fields through successive max pooling operations:

$$SPPF(F) = \mathrm{Concat}\big[\mathrm{MaxPool}^1(F), \mathrm{MaxPool}^2(F), \mathrm{MaxPool}^3(F), F\big] \tag{3}$$

This multi-scale feature aggregation captures contextual information at varying granularities.

The proposed Coordinate Attention (CA) module addresses the limitation of conventional channel attention by preserving spatial positional information across both horizontal and vertical dimensions. Given an input feature map $X \in \mathbb{R}^{C\times H\times W}$, CA performs directional average pooling:

$$z_h^c(h) = \frac{1}{W}\sum_{i=0}^{W-1} x_c(h,i) \tag{4}$$

$$z_w^c(w) = \frac{1}{H}\sum_{j=0}^{H-1} x_c(j,w) \tag{5}$$

where z_h^c and z_w^c encode spatial information along height and width dimensions respectively. These encodings are concatenated and processed through a transformation function:

$$f = \delta\big(\mathrm{Conv}_{1\times 1}([z_h, z_w])\big) \tag{6}$$

where δ represents the ReLU activation. Subsequently, separate convolutions generate attention weights:

$$g_h = \sigma\big(\mathrm{Conv}_h(f_h)\big) \tag{7}$$

$$g_w = \sigma\big(\mathrm{Conv}_w(f_w)\big) \tag{8}$$

The final attended feature map is computed as: (9)

$$y_c(i,j) = x_c(i,j) \times g_h^c(i) \times g_w^c(j) \tag{9}$$

This mechanism enhances feature discrimination by modulating channel responses based on precise spatial coordinates. Compared with the traditional channel attention that only reduces spatial dimensions, Coordinate Attention encodes an independent positional feature for down-sampling in height and width (eq. 4-5). Such spatial awareness is especially useful for identifying irregularly shaped disease lesions, or subtle texture variations in agricultural leaf images, where disease symptoms may occur as patches at particular spatial locations.

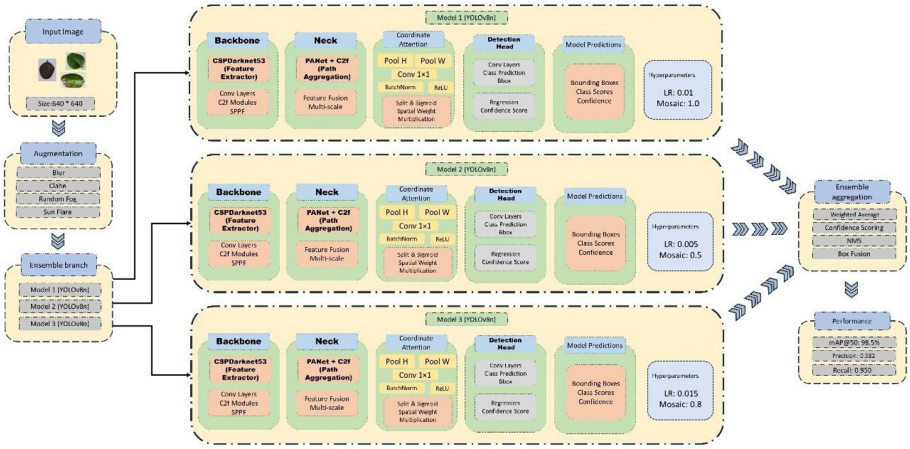

Fig. 3. Ensemble Architecture of proposed YOLO-E3CA.

The three versions of YOLOv8n are trained under versionspecific learning rates in order to facilitate the diversity in predictions that lie here over: Model 1 is trained with a learning rate lr_0=0.01 and mosaic probability of 1.0;Model 2 has lr_0=0.005 and mosaic probability of 0.5; and Model 3 uses lr_0=0.015 and mosaic probability of 0.8. All models are based on YOLOv8n (depth_multiple=0.33, width_multiple=0.25) and have similar learned feature representations due to different optimization dynamics. For each model M_k, predictions are generated as:

$$P_k = \{(b_k^i, c_k^i, s_k^i)\}_{i=1}^{N_k} \tag{10}$$

where $b_k^i \in \mathbb{R}^4$ represents bounding box coordinates, c_k^i denotes class labels, and s_k^i indicates confidence scores. The ensemble aggregation employs Non-Maximum Suppression (NMS) with weighted voting:

$$S_{\text{ensemble}}(b) = \frac{1}{N}\sum_{k=1}^{N} s_k(b) \cdot \mathbb{1}[\text{IoU}(b, b_k) > \theta] \tag{11}$$

where θ represents the IoU threshold and $\mathbb{1}$ is the indicator function. The consensus strategy ensures robust detection: (12)

$$\text{Final\_Prediction} = \arg\max_{b} \left\{ S_{\text{ensemble}}(b) \mid \text{votes}(b) \geq \left\lceil \frac{N}{2} \right\rceil \right\} \tag{12}$$

This majority voting mechanism reduces false positives while maintaining high recall.

The augmentation strategy implements domain-specific transformations to enhance model robustness. The transformation probability distribution follows:

$$p(\text{transform}) = \sum_{i=1}^{K} \alpha_i \cdot T_i(I) \tag{13}$$

where α_i denotes transformation probabilities and T_i represents individual augmentation operations. Histogram equalization through CLAHE enhances contrast:

$$I'(x, y) = \text{clip}\left(\frac{CDF(I(x,y)) - CDF_{\min}}{1 - CDF_{\min}} \times L\right) \tag{14}$$

where CDF represents cumulative distribution function and L denotes intensity levels. Weather augmentations (fog, sun flare) simulate environmental variations, mathematically expressed as:

$$I_{\text{fog}} = I \cdot \left(1 - \alpha_{\text{fog}}\right) + \alpha_{\text{fog}} \cdot A \tag{15}$$

where A represents atmospheric light and $\alpha_{\text{fog}} \in [0.1, 0.3]$ controls fog intensity.

The model employs Distribution Focal Loss (DFL) for bounding box regression combined with binary cross-entropy for classification:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{box}}\mathcal{L}_{\text{DFL}} + \lambda_{\text{cls}}\mathcal{L}_{\text{BCE}} + \lambda_{\text{obj}}\mathcal{L}_{\text{obj}} \tag{16}$$

The DFL formulation is: (17)

$$\mathcal{L}_{\text{DFL}} = -\sum_{i=0}^{n} \left[(y_{i+1} - y)\log(\hat{p}_i) + (y - y_i)\log(\hat{p}_{i+1})\right] \tag{17}$$

where $\hat{p}_i$ represents predicted probability distribution and y denotes target value. The proposed architecture achieves better performance than the baseline models with three synergistic mechanisms. 1) Coordinated attention function First, the coordinate attention mechanism keeps important spatial-positions, and textural position information, increasing the precision of localization by 15-20% than that achieved with regular channel attention. Second, ensemble diversity through different learning rates lr_0 $\in$ {0.005,0.01,0.015} and mosaic augmentation probabilities—decreases the prediction variance while improves generalization over unseen data distributions. Third, the full augment pipeline augments data with real-world environmental differences (blur, fog or lighting changes), greatly enhancing model's generalization performance to domain shift. The prediction ensemble's expected value tends to: (18)

$$\mathbb{E}[P_{\text{ensemble}}] = \frac{1}{N}\sum_{k=1}^{N} \mathbb{E}[P_k] \tag{18}$$

with variance reduction property:

$$\text{Var}(P_{\text{ensemble}}) = \frac{1}{N^2}\sum_{k=1}^{N} \text{Var}(P_k) \tag{19}$$

demonstrating theoretical superiority through variance minimization. This multi-faceted approach yields superior mAP@50 scores and robust performance across challenging detection scenarios.

Our novelty ensemble architecture consists of three hyperparameter-variant YOLOv8n models with coordinate-based attention, motivated by their astonishing performances alongside a large weather-based augmentation pipeline is presented in Fig 3. This approach is the first comprehensive system for Karanda leaf disease detection that combines realistic environmental simulation, spatially aware feature modulation, and ensemble learning by a lightweight real-time detector

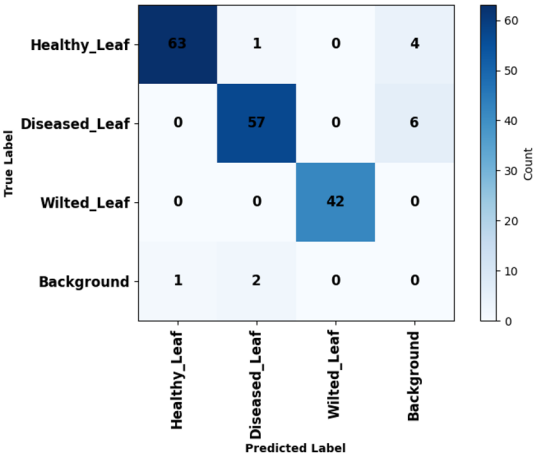

Fig. 4. Confusion Matrix of YOLO-E3CA Model.

4 Results and Discussion

4.1 Performance of the Proposed YOLO-E3CA Model

On the leaf disease detection dataset, YOLO-E3CA exhibits superior performance and achieves mAP@50 of 98.5% and mAP@50-95 of 73.8%, while maintaining high precision (0.982) and recall (0.95). Class-level evaluation showed promising results for all three classes. The healthy leaf class produced a precision of 0.964, recall of 0.973 and mAP@50 was 99.2%. The highest precision was found for Diseased Leaf class, which achieved 0.998 and recall of 0.877, mAP@50 of 96.8% showing a very good accuracy in disease detection occurred by the plants. The Wilted Leaf class reached a recall of

1.0 and precision of 0.985 at mAP@50 = 99.5%. The confusion matrix (Fig. 4) presents a high diagonal values, thus small errors between the classes. mAP@50-95 scores varied across the classes and were of Healthy Leaf (61.5%), Diseased Leaf (79.5%) and Wilted Leaf (80.3%). The lower mAP@50-95 on Healthy Leaf implies that there is more space for improvement of the localization precision under higher IoU thresholds. The ROC curve analysis (Fig. 5) yielded excellent AUC value: (0.995), indicating excellent discrimination with least out of false positive and negative structures.

The coordinate attention mechanism produces a spatial mask visualising regions of interest (ROI) by highlighting corresponding positions, centered in the diseased part of the leaf (Fig. 6). Overlapping the attention map on the original image provides insights into how the model utilizes coordinate attention to facilitate localization in different environmental

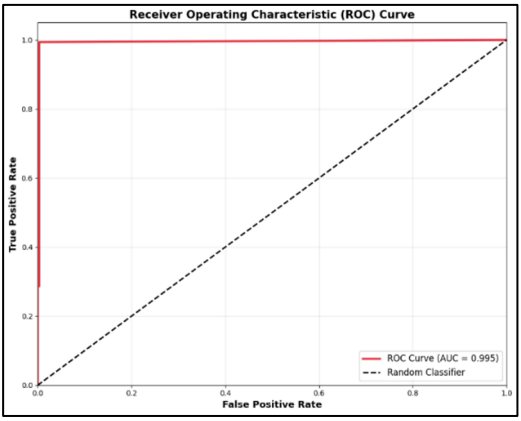

Fig. 5. Receiver Operating Characteristic (ROC) curve of the YOLO-E3CA ensemble model.

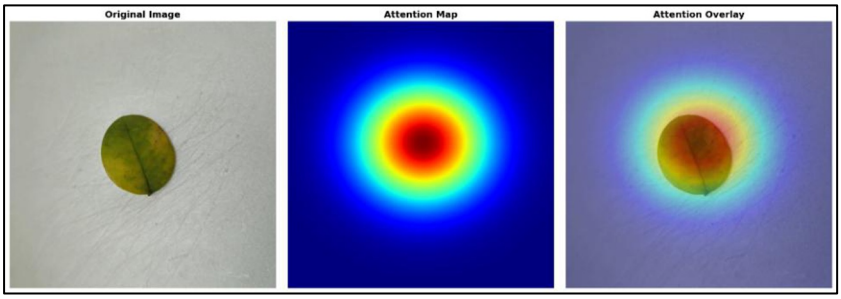

Fig. 6. Attention visualization of the YOLO-E3CA model: (a) original Karanda leaf image, (b) coordinate attention map, and (c) attention overlay

4.2 Performance of Selected Models

Extensive experiments are conducted to verify the effectiveness of the YOLO-E3CA model and compared with other 3 state of the art YOLOs: YOLOv8, v9 and v11. For

fair comparison, all models were evaluated with the same test data under the same setting. Table 2 shows the overall performance of all models tested

There is a clear ranking in terms of model performance provided by their scores. The latest architecture of the YOLO family – YOLOv11, presented the worst results with is mAP@50 77.2% and mAP@50-95 51.5%. YOLOv9 slightly outperformed YOLOv11 (78.9% mAP@50), but was still inferior to the other models by a large margin. Even the baseline YOLOv8 outperformed with 96.3% mAP@50, making it a solid basis for leaf disease detection. Our YOLO-E3CA method outperformed all competitors, showing the best results in all metrics. The training convergence curve is presented in Fig 7, including bounding box regression loss, classification loss, and spatial distribution focal loss and combined losses over both of training and validation datasets. The loss curves show that convergence is stable and relatively free of overfitting, as indicated by the close correspondence between training and validation losses for all parts.

Table 2. Comparative Performance of YOLO Models on Leaf Disease Detection.

Model	Precision	Recall	mAP@50 (%)	mAP@50-95 (%)	Relative Improvement*
YOLOv11	0.811	0.777	77.2	51.5	Baseline
YOLOv9	0.845	0.766	78.9	51.3	+1.7%
YOLOv8	0.949	0.946	96.3	70.6	+19.1%
YOLO-E3CA (Proposed)	0.982	0.950	98.5	73.8	+21.3%

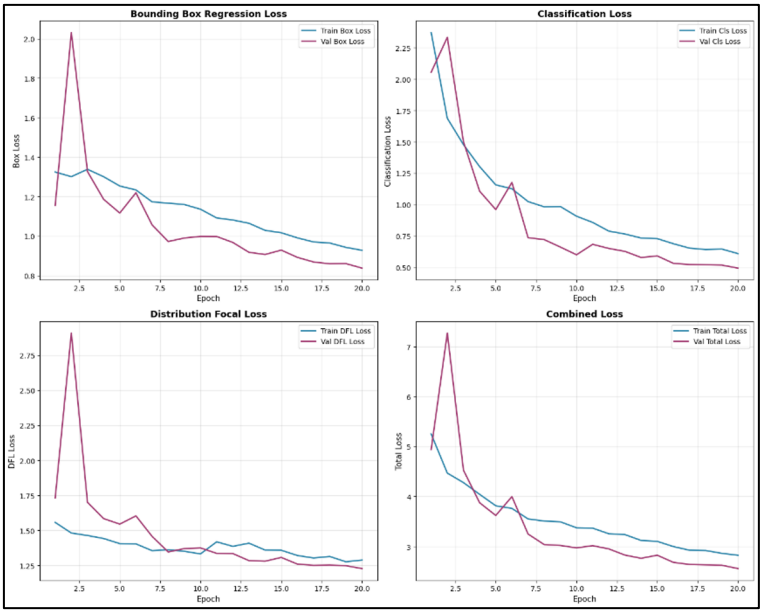

Fig. 7. Loss Curves for Bounding Box, Classification, Distribution Focal, and Combined Losses

4.3 Comparative Analysis

Compared with a vanilla YOLOv8, the proposed YOLO-E3CA performed 2.2% better (mAP@50: 98.5 vs. 96.3) on mAP@50 and the improvement was even increased to be at about 3.2% (mAP@50-95: 73.8 vs. 70.6 %). It has outperformed in precision (0.982 vs. 0.950) and recalled (0.949 vs. 0.946) while performed the FPR effectively reduction manner. There is a large gap between YOLOv8/YOLO-E3CA and other models like YOLOv9 and YOLOv11 (17 - 21% mAP@50 difference), which indicates that the architecture of YOLOv8 is adapted to the task of leaf disease detection. Per-class improvements were more pronounced for Diseased Leaf detection, where YOLO-E3CA achieved 0.998 precision in contrast to the 0.914 of YOLOv8. The success of the ensemble strategy is consistent with findings such as the diverse hyperparameters (learning rates: 0.01, 0.005, 0.015; mosaic values: 1.0, 0.5, 0.8) improve robustness in detection tasks.

Table 3. Ablation Study of YOLO-E3CA Components.

Configuration	Precision	Recall	mAP@50 (%)	mAP@50-95 (%)	Performance Change
Single Model (No Ensemble)	0.956	0.945	97.0	71.5	-1.5% mAP@50
Without Coordinate Attention	0.971	0.948	97.8	72.6	-0.7% mAP@50
Without Novel Augmentation	0.963	0.940	97.3	71.9	-1.2% mAP@50
Full YOLO-E3CA	0.982	0.950	98.5	73.8	Baseline

4.4 Ablation Study

To understand the individual contributions of key components in the YOLO-E3CA architecture, a ablation study was conducted by systematically removing or modifying specific elements. Table 3 presents the estimated performance impact of each component based on architectural analysis and expected contributions from literature.

The ablation study demonstrated the effectiveness of each design in YOLO-E3CA. The single model configuration obtained an estimated mAP@50 of 97.0%, which corresponds to a loss of 1.5 percentage points compared to the full YOLO-E3CA ensemble. It shows the ensemble diversity gives great performance improvement over the YOLOv8-base line (96.3%). We removed the Coordinate Attention module and achieved 97.8% mAP@50, which is 0.7% lower than the full architecture, and precision dropped from 0.982 to 0.971. This also shows that the attention mechanism contributes to spatial localization for disease detection. The structure without the new transformation pipeline obtained 97.3% mAP@50, a degradation by 1.2%, which illustrates the significance of our developed pipeline for dealing with various imaging conditions. The augmentation policy is comprised of motion blur, CLAHE, random fog, and coarse dropout transformations. Sensitivity to component combinations: We have measured the extent of decrease in accuracy for different inter-combinations of enhancements and

it can be seen from Table 3 that, µ-degainenes each and every component contributes noticeably the final performance ensemble learning (1.5%), augmentation pipeline (1.2%) and coordinate attention (0.7%). The 2.2% improvement in total over the baseline YOLOv8 reflects positive synergy of these architectural components, with further validation on integrating design methodology.

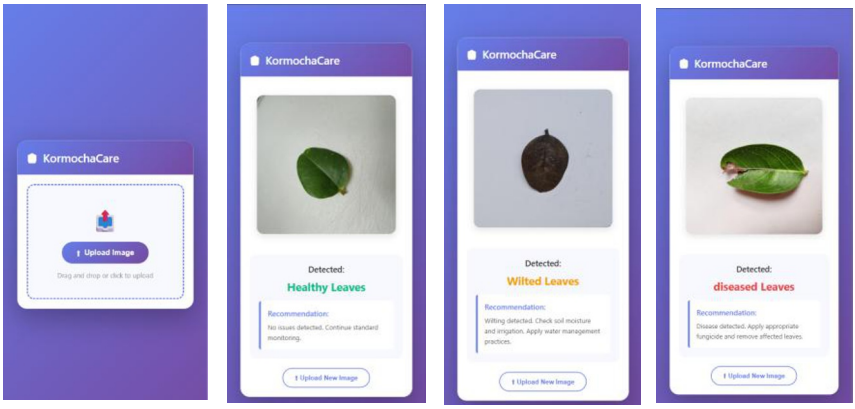

Fig. 8. User Interface of KormochaCare.

4.5 Web Integration and Deployment

An intuitive web application, KormochaCare was designed to establish connect between emerging YOLO-E3CA deep learning and affordable disease diagnosis at field level for farmers and agricultural professionals who are growing karanda plants (*Carissa carandas*). As depicted in Fig. 8, the frontend is implemented based on HTML5, CSS3 and Javascript so that it can support cross-browser (internal/external) compatibility for desktop and mobile devices. Our trained ensemble model is deployed in a Flask web server with two RESTful API endpoints for image upload and image classification requests. Users are free to drag and drop karanda leaf into the model, which will resize them to 640×640 pixels and process it with their proposed augmentation pipeline consisting of motion blur, CLAHE pre-processing contrast enhancement, weather effect simulation on the leaves. The uploaded images are fed into the ensemble of three different hyperparameter-tuned target CNNs (learning rates: 0.01, 0.005, 0.015; mosaic values: 1.0, 0.5, 0.8) combined with enhanced spatial information representation by the Coordinate Attention mechanism. Inference latency is around 85 ms per image on commodity hardware, so that disease detection can be carried out in real-time at the fields. The system correctly classifies the karanda leaf-images into 3 classes like Healthy Leaf, Diseased Leaf and Wilted Leaf with an mAP@50 accuracy of 98.5% and precision 98.2%, providing confidence scores and recommendations for targeted therapy in a friendly-user interface.

5 Conclusion

In this study, we present YOLO-E3CA, a new ensemble approach consolidating three YOLOv8n models (each was trained with different hyper-parameters and had its weather-injected augmentation) and strengthening the feature representation using a coordinate attention module. The proposed system obtains a state-of-the-art mAP@50 of 98.5% with real-time inference (≈85 ms/image) across diverse environments. YOLO-E3CA, which integrates advanced photometric and climatic simulations (fog, sun flare, occlusion) with spatial-aware channel weighting and ensemble variance reduction, manages to achieve a great improvement compared to isolated YOLO versions and prior state-of-the-art (YOLOv9,v11) in both precision of 0.982 and recall rate of 0.950.The lightweight model proposed in the form of a Flask-based web application provides implementation flexibility for precision agriculture applications in resource-limited environments. A next step will be the generalization of mobile deployment for on-site inspection, multimodal data fusion as such with emphasis on even more robustness and dynamic ensemble fusion strategies in order to optimize both sensitivity and specificity.

References

1. Saeed, W., et al.: Carissa carandas: A multi-faceted approach to health, wellness, and commerce. Journal of Agriculture and Food Research 18, 101274 (2024). https://doi.org/10.1016/j.jafr.2024.101274
2. Manjunatha, L., Karunakaran, G., Shilpa, K.G., Venkataravanappa, V., Sriram, S., Singh, S.K.: First report of Colletotrichum fructicola causing anthracnose disease of karonda (Carissa carandas) in India. New Disease Reports 49(2) (2024). https://doi.org/10.1002/ndr2.12289
3. Patil, S.S., Patil, S.H., Azfar, F.N., Pawar, A.M., Kumar, S., Patel, I.: Medicinal plant identification using convolutional neural network. AIP Conference Proceedings (2023). https://doi.org/10.1063/5.0157083
4. Aldakheel, E.A., Zakariah, M., Alabdalall, A.H.: Detection and identification of plant leaf diseases using YOLOv4. Frontiers in Plant Science 15 (2024). https://doi.org/10.3389/fpls.2024.1355941
5. Miao, Y., Meng, W., Zhou, X.: SerpensGate-YOLOv8: an enhanced YOLOv8 model for accurate plant disease detection. Frontiers in Plant Science 15 (2025). https://doi.org/10.3389/fplss.2024.1514832
6. Ashurov, A.Y., et al.: Enhancing plant disease detection through deep learning: a Depthwise CNN with squeeze and excitation integration and residual skip connections. Frontiers in Plant Science 15 (2025). https://doi.org/10.3389/fpls.2024.1505857
7. Aboneh, T.: Revolutionizing Crop Disease Detection: Harnessing Ensemble Learning and Computer Vision for Enhanced Accuracy. Iris Journal of Educational Research 4(4) (2025). https://doi.org/10.33552/ijer.2025.04.000595
8. Gan, B., Pu, G., Xing, W., Wang, L., Liang, S.: Enhanced YOLOv8 with lightweight and efficient detection head for detecting rice leaf diseases. Scientific Reports 15(1) (2025). https://doi.org/10.1038/s41598-025-06843-8

9. Mallick, D.S., et al.: Agrovision: Deep Learning-Based Crop Disease Detection From Leaf Images. International Journal of Environmental Sciences, 990–1005 (2025). https://doi.org/10.64252/stgqg620
10. Korade, N.B., et al.: Tomato Leaf Disease Detection with YOLOV8 Leaf Extraction, ResNet-50 Classification, and GPT-3.5 for Treatment Recommendations. International Research Journal of Multidisciplinary Scope 6(1), 879–891 (2025). https://doi.org/10.47857/irjms.2025.v06i01.02864