Deep Learning-Based Pomegranate Growth Stage Detection Using YOLOv11 for Precision Agriculture

1st Gopi M
gopim0504@gmail.com
Computer science and engineering
UVCE, Bengaluru-560001, Karnataka, India

2nd Kumaraswamy
kumaraswamy.shivashankar@gmail.com
Computer science and engineering
UVCE, Bengaluru-560001, Karnataka, India

***Abstract*—Pomegranate cultivation in India faces challenges in growth stage monitoring, impacting yield and resource efficiency. This paper proposes a YOLOv11-based model to automate pomegranate growth stage detection (bud, flower, early fruit, mid-growth, ripe) using high-resolution images. The model achieves a mean average precision (mAP@50) of 0.875 and mAP@50-95 of 0.723, outperforming manual methods. Trained on 5,858 annotated images (80% training, 10% validation/testing), the system integrates data augmentation and hyperparameter tuning for robustness. Results demonstrate its potential for precision agriculture, enabling optimized irrigation, pest control, and harvesting. Experimental results show that the model achieves mAP@0.5 of 87.5% and mAP@0.5–0.95 of 72.3%, with high real-time inference capability. Visual analysis through confidence-threshold and precision-recall curves further validates the model's robustness. This system offers a scalable, fast, and reliable solution for real-time monitoring and decision-making in pomegranate farming. Its integration with IoT-based platforms can significantly aid farmers in optimizing irrigation, fertilization, pest control, and harvesting schedules—ultimately improving yield quality and reducing losses.**

***Index Terms*—Pomegranate growth stages, YOLOv11, precision agriculture, deep learning, object detection**

I. Introduction

Pomegranate (Punica granatum) is one of the most economically significant fruit crops cultivated across India, especially in semi-arid regions such as Maharashtra, Karnataka, and Andhra Pradesh. Known for its nutritional richness—being an excellent source of vitamin C, iron, and antioxidants—pomegranate also holds cultural, medicinal, and industrial value. Its consumption spans fresh fruit, juice extraction, and pharmaceutical and dye industries. Despite its agricultural importance, effective monitoring of pomegranate crop development remains a critical challenge, especially in large-scale farms.

The growth of pomegranates occurs through several distinct stages: bud formation, flowering, early fruit, mid-growth, and ripening. Each stage is crucial, not only for yield prediction and harvesting but also for effective irrigation, nutrient management, and pest control. Traditional methods of monitoring crop growth rely heavily on manual labor and visual inspection. These methods are time-consuming, subjective, and prone to errors, particularly in dynamic environmental conditions or when monitoring vast orchards. Misidentifying growth stages can result in improper pesticide usage, suboptimal irrigation schedules, and premature or delayed harvesting, directly affecting fruit quality and overall yield.

With the increasing demand for sustainable and intelligent farming solutions, the integration of Artificial Intelligence (AI), particularly deep learning and computer vision, offers promising avenues for automating critical agricultural tasks. Deep learning has revolutionized visual recognition tasks by enabling machines to automatically learn and detect complex visual features from images. Among various deep learning architectures, the YOLO (You Only Look Once) family has emerged as one of the most efficient frameworks for real-time object detection due to its speed, accuracy, and end-to-end learning capability.

This paper introduces a novel application of YOLOv11—a state-of-the-art object detection model—for identifying the growth stages of pomegranates from high-resolution images. YOLOv11 builds upon its predecessors by incorporating architectural improvements such as the Spatial Pyramid Pooling-Fast (SPPF) layer for better spatial feature representation and the C2PSA attention block for enhanced object localization. These enhancements are especially beneficial for agricultural images, where objects such as buds and small fruits can vary significantly in size, color, and texture.

In this study, a custom image dataset of 5,758 images was developed and labeled into five major growth categories. The model was trained using 80% of the dataset, validated on 10%, and tested on the remaining 10%. A series of data augmentation techniques were employed to enhance the robustness of the model, including flipping, scaling, and color normalization. The performance of the trained model was evaluated using standard metrics such as mean Average Precision (mAP), precision, recall, and F1-score.

The objectives of our work are,

1) By accurately classifying pomegranate growth stages, the system provides actionable insights to farmers, enabling them to take timely decisions that improve fruit quality and optimize resource use.

2) For instance, the detection of the flowering stage could prompt targeted pollination strategies, while the identification of the ripening stage could trigger precision harvesting and reduce post-harvest losses.
3) Given its low computational complexity and high detection accuracy, YOLOv11 is suitable for mobile and drone-based applications, thereby offering a practical pathway toward fully automated, intelligent farm monitoring systems.

The rest of the sections are organized as follows: Section II gives a detailed discussion on the existing literature and identifies the research gaps from it. Section III examines the custom dataset used in this experimentation. Section IV gives insights on the various YOLO architectures used in this experimentation, and Section V discussed proposed methodology of our work. Section VI and VII analyze the results obtained through experimentation and provide a conclusion of our work.

II. Background

The following section examines the recent literature from SCOPUS and IEEE databases. Initially, 25 papers were shortlisted for the review based on the keyword searches of ”grape leaf disease identification” and ”grape leaf disease classification” from the aforementioned databases. Then the papers were sorted based on year, relevance, and indexing, and then 10 papers were selected for the review.

The work at [1] comparative study of object detection models—YOLOv8, YOLOv9, YOLOv10, YOLOv11, and Faster R-CNN—for detecting multiple weed species in agricultural fields. The study addresses the challenge of site- and species-specific weed management, especially under increasing herbicide resistance. Researchers developed an annotated image dataset of five weed species and trained the models on this data. Among the models, YOLOv9 achieved the highest detection accuracy (mAP@0.5 of 0.935), while YOLOv11 demonstrated the fastest inference time (13.5 ms). YOLOv8 and YOLOv10 also balanced accuracy and speed effectively, outperforming the two-stage Faster R-CNN, which had a slower inference time (63.8 ms) and lower accuracy (mAP@0.5 of 0.821).

The study at [2] explores the application of the YOLOv11 object detection model for identifying polyps in colonoscopy images, aiming to support early colorectal cancer detection. The authors compare five variants of YOLOv11—n, s, m, l, and x—using the Kvasir dataset, both in its original form and an augmented version. YOLOv11, building on YOLOv8, introduces architectural improvements like the C2PSA and C3K2 modules for better precision and efficiency. Results showed that the lightweight YOLO11n model offered the best balance of precision and F1-score relative to its low parameter count, especially after augmentation.

The paper in [3] introduces a novel framework for generating remote sensing (RS) images from spatial relationship descriptions using a two-stage pipeline. First, a semantic structuring model transforms spatial text descriptions into structured layouts, capturing object relationships, directions, and distances. Then, an enhanced diffusion model called GeoRSDiffusion synthesizes the final image using positional prompts and a layout attention mechanism to maintain spatial fidelity. The model was trained and evaluated using a custom dataset, RS5layout, covering five geographic object categories. Experiments showed that GeoRSDiffusion significantly outperforms existing methods like LayoutDiffusion, GLIGEN, and ALDM in both image quality and spatial accuracy.

The paper in [4] a comprehensive analysis of the YOLO (You Only Look Once) series from YOLOv1 through YOLOv11, highlighting architectural changes, performance improvements, and application domains. YOLO started as a real-time object detection model that predicts bounding boxes and class probabilities in a single forward pass, offering a faster alternative to two-stage detectors like R-CNN. Over time, each version introduced significant upgrades—YOLOv2 brought anchor boxes, YOLOv3 added multi-scale prediction, and YOLOv4 implemented CSPDarknet and enhanced training strategies.

The work at [5] introduces CRFUSION, a multimodal object identification system that combines RGB images from cameras and mmWave radar signals to classify both the category and texture of objects with high precision. To extract meaningful features from the RF signal, the authors propose a novel metric called the Energy Reflection Factor (ERF), which captures both shape and texture information. CRFUSION uses a dual-input neural network (CRFNET) that fuses ERF features and image embeddings using a multi-head attention mechanism for accurate classification. It was evaluated using a custom dataset of 16 everyday objects across six categories and seven textures, achieving over 94

This paper reviews in [6] a real-time Automatic License Plate Recognition (ALPR) system tailored for Moroccan license plates using the YOLOv3 deep learning model. The system operates in three main stages: vehicle detection, vehicle tracking using the DeepSort algorithm, and license plate detection and recognition. A voting mechanism aggregates character recognition results across frames to improve accuracy. The authors collected a custom dataset featuring diverse environmental conditions and Arabic characters, which was used to train and evaluate the system. The model achieved high accuracy: 99% for vehicle detection, 99% for license plate detection, and 94.5% for character recognition. YOLOv3 was chosen for its efficiency on edge devices like the NVIDIA Jetson AGX.

The paper at [7] GFS-YOLO11, a lightweight and accurate tomato maturity detection model tailored for both common and cherry tomatoes in complex field conditions. It builds upon YOLO11 by introducing three key modules: C3k2_Ghost for reducing computation, FRM (Feature Refining Module) to recover feature expressiveness, and SPPFELAN for multi-scale feature fusion. A custom dataset called Tomato-Detect, containing six maturity categories, was used for training and evaluation. The model achieved superior performance, reaching 92% precision, 86.8% recall, and 93.4% mAP@0.5, while also reducing parameters and inference time significantly.

The paper at [8] an adaptive YOLO11-based framework designed for detecting, tracking, and imaging small aerial targets using a pan–tilt–zoom (PTZ) camera network. The system integrates stereo vision and deep learning, offering a cost-effective alternative to radar for real-time surveillance. To enhance small object detection, the authors propose advanced data augmentation techniques using SAM, Stable Diffusion, and GANs, along with knowledge distillation. YOLOv11x achieved the highest precision (mAP50 of 86.7%) among tested models, while YOLOv8n offered the fastest inference at 0.6 ms.

The paper at [9] investigates the use of YOLOv9, YOLOv10, and YOLOv11 algorithms to detect various defects in solar panels using thermal and optical images. The study compares performance across three datasets and finds that YOLOv11-X delivers the best results, achieving a high F1 score of 90% and mean average precision of 92.7%. Key innovations in YOLOv10 and YOLOv11, such as NMS-free training and the C2PSA block, enhance detection speed and accuracy. Experimental results show that YOLOv10-X and YOLOv11-X outperform traditional methods like SVM and Faster R-CNN in both precision and inference efficiency.

The work at [10] introduces YOLO-E, a lightweight object detection model optimized for military target detection under resource-constrained environments like drones. Built upon YOLOv8n, YOLO-E incorporates GhostConv, EMSC modules, a shared convolutional detection head, and a new bounding box loss function—NCDIoU—to improve accuracy and efficiency. The authors created a custom dataset of 7,347 images with annotated military targets for evaluation. YOLO-E achieves a 2.33% accuracy improvement over YOLOv8n, while reducing parameters by 30.87% and computation by 37.33%.

The following are the research gaps identified through this literature review,

1) Lack of comparative studies on the performance of YOLO architectures
2) Limited focus on integrating advanced attention mechanisms with YOLO models for enhanced feature extraction and real-time efficiency
3) Insufficient exploration of augmented datasets tailored specifically for grape leaf diseases to address class imbalance and dataset diversity

III. Dataset Description

The dataset used in this study consists of a total of 5,758 high-resolution images of pomegranates, systematically categorized into five distinct growth stages: bud, flower, early-fruit, mid-growth, and ripe. Each image is meticulously labeled and annotated to reflect its respective category, enabling accurate training of machine learning models. The dataset was preprocessed by resizing the images to 640×640 pixels and normalizing the pixel values. It was then split into 80% for training (4,606 images), 10% for validation (578 images), and 10% for testing (578 images). To further enhance model generalization, data augmentation techniques such as rotation, flipping, and scaling were employed, expanding the dataset's diversity and helping the model adapt to real-world variations. The annotations and image quality play a critical role in enabling precise classification of pomegranate growth stages using deep learning models, particularly the YOLOv11 architecture. This robust and balanced dataset forms the foundation for developing an effective and scalable agricultural monitoring system that supports precision farming and decision-making.

The dataset used for experimentation is a custom dataset of grape leaves collected from a Local farm consisting of 200 images. The 200 image samples are a combination of "Fresh" and "Diseased" grape leaves with the use case of disease identification and classification. Each of the images were of 2080 x 4608 pixel size with a bit depth of 24. The dataset is augmented to 500 images making it suitable for the analysis using YOLO architectures. The figure 1 shows a sample dataset collected for this experimentation. Table I summarizes the dataset statistics.

IV. About Learners

To do experimentation, we have used YOLO architecture YOLOv11. The following section gives a brief discussion on the specialties of architectures,

A. YOLOv11

YOLOv11 focuses on small object detection and introduces the Spatial Pyramid Pooling Fast (SPPF) module and advanced attention mechanisms like the C2PSA block. The loss function emphasizes small object detection:

$$L = L_{\text{loc}} + \gamma \times L_{\text{small\_obj}} + L_{\text{cls}} \quad (1)$$

where:

- L_{loc}: Localization loss
- $L_{\text{small\_obj}}$: Loss component for small objects
- L_{cls}: Classification loss
- γ: Weighting factor for small object loss

The table below provides a YOLOv11, highlighting their key features and innovations.

V. Proposed Methodology

Figure 3 shows our proposed methodology. The dataset was collected and preprocessed to make it suitable for YOLO architectures. The dataset was segregated into train, validation, and test sets in the ratio of 70, 20, and 10% respectively. The dataset was trained on the YOLO v11 architecture independently, and its performance is validated using the test dataset. The performance of the model is recorded and compared against each other to infer the highest-performing model.

The fig 3 presents a structured workflow for developing and deploying a YOLOv11-based deep learning model to identify the growth stages of pomegranates. The process begins with data collection, where a dataset of 5,857 labeled images is gathered, representing five distinct growth stages: bud, flower, early fruit, mid-growth, and ripe. This comprehensive dataset

Fig. 1: Dataset Samples, Image(1)-Bud, Image(2)-flower, Image(3)-Early-fruit, Image(4)-Mid-Growth, Image(5)-Ripe

TABLE I: Pomegranate Dataset Description

Feature	Description
Number of Original Samples	5855 images
Categories	Bud, Flower, Mid-Growth, Early-Fruit and Ripe
Use Case	Growth stage identification and classification
Image Dimensions	640 x 480 pixels
Bit Depth	24-bit
Augmented Dataset Size	5855 images
Analysis Framework	Suitable for YOLO architectures

TABLE II: Summary of YOLOv11

Feature	YOLOv11
Backbone	Advanced Convolutional Blocks
Neck	Spatial Pyramid Pooling Fast (SPPF)
Detection Head	Anchor-Free
Special Focus	S Small Object Detection with Attention Mechanisms
Loss Function	Small Object Weighted Loss: $L = L_{\text{loc}} + \gamma \times L_{\text{small obj}} + L_{\text{cls}}$

provides the foundation for training an accurate and robust detection system.

The Algorithm 1 shows the overall flow of our work.The algorithm for grape leaf disease identification and classification involves several structured steps leveraging YOLO architectures (YOLOv11). Initially, the custom dataset (D) comprising 5787 labeled images of grape leaves (”bud” and ”flower” and ”early-fruit” and ”mid-growt” and ”ripe”) is augmented to expand the dataset size to 6000 images. Preprocessing includes resizing images to 640×640 pixels and normalizing pixel values ($I' = I/255, \forall I \in \mathsf{D}$). The dataset is then split into training ($\mathsf{D}_{\text{train}}$, 70%), validation ($\mathsf{D}_{\text{val}}$, 15%), and testing ($\mathsf{D}_{\text{test}}$, 15%) subsets. YOLO model M_x, where $x \in \{11\}$, is initialized with pre-trained weights and trained on $\mathsf{D}_{\text{train}}$ using a composite loss function ($\mathsf{L} = \mathsf{L}_{\text{loc}} + \mathsf{L}_{\text{conf}} + \mathsf{L}_{\text{cls}}$). Performance is evaluated on D_{test} based on metrics such as mean Average Precision (mAP_x), precision, recall, F1-score, and inference speed (FPS_x). Hyperparameters (α, β, γ) are fine-tuned for each model using D_{val}, and training is repeated for optimization. The model with the best performance ($\mathsf{M}^* = \arg\max_{\mathsf{M}_x} \text{mAP}_x$) is selected for deployment and real-time classification, with potential integration into IoT systems for enhanced growth stage monitoring. This systematic approach ensures efficient training, evaluation, and deployment of an optimal YOLO architecture for pomegranate growth stages and classification.

VI. Results and Discussion

The fig 3 presents training and validation performance metrics for a YOLOv11 model, likely trained for object detection, such as pomegranate growth stage classification. The top row shows training metrics: box loss, classification loss, distribution focal loss (DFL), precision, and recall. These losses steadily decrease, indicating effective learning and convergence. Meanwhile, precision and recall steadily improve—precision reaching above 0.9 and recall rising past 0.85—suggesting that the model becomes more accurate and less likely to miss detections over epochs. The smoothing

Algorithm 1 Pomegranate Growth Stages Identification and Classification using YOLOv11

Let the dataset be defined as

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where x_i is a high-resolution image of a pomegranate and $y_i \in$ {bud, flower, early-fruit, mid-growth, ripe} is the corresponding label for the growth stage. **Step 1: Data Augmentation and Preprocessing.** Apply data augmentation on the dataset D using transformations such as horizontal flipping, rotation, and scaling to create a more diverse training set D_{aug}. Normalize each image using:

$$x_i' = \frac{x_i}{255}$$

Resize all images to a fixed dimension of $640 \times 640 \times 3$ pixels. **Step 2: Dataset Splitting.** Divide the augmented dataset into three subsets:

$$D = D_{\text{train}} \cup D_{\text{val}} \cup D_{\text{test}}$$

with 70% for training, 15% for validation, and 15% for testing. **Step 3: Model Initialization.** Initialize the YOLOv11 model M with pre-trained weights ϑ_0. Modify the output layer to predict five classes corresponding to the pomegranate growth stages. **Step 4: Loss Function Definition.** Use a composite loss function to train the model:

$$L = L_{\text{box}} + L_{\text{conf}} + L_{\text{cls}}$$

where L_{box} is the bounding box regression loss, L_{conf} is the objectness confidence loss, and L_{cls} is the classification loss. **Step 5: Model Training.** Train the model on batches from D_{train} using gradient descent to update weights:

$$\vartheta \leftarrow \vartheta - \alpha \nabla_{\vartheta} L$$

where α is the learning rate. **Step 6: Model Evaluation.** Evaluate the model performance on the validation set using precision, recall, and F1-score:

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}, \quad F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Also compute mean Average Precision at IoU thresholds:

$$\text{mAP}_{@0.5}, \quad \text{mAP}_{@[0.5:0.95]}$$

Step 7: Hyperparameter Tuning. Adjust hyperparameters such as learning rate α, batch size b, and epochs e using grid search or cross-validation to optimize model performance. **Step 8: Model Selection.** Choose the best-performing model as:

$$M^* = \arg\max_{M} \text{mAP}_{\text{val}}$$

Step 9: Deployment. Deploy the final model M^* for real-time detection. For a new input image x, the model produces a predicted label:

$$\hat{y} = M^*(x), \qquad \hat{y} \in \mathbb{R}^5$$

The output also includes bounding box coordinates (x, y, w, h) and a confidence score for each detected object.

lines (orange dashed) help visualize overall trends despite some fluctuations. The bottom row shows validation metrics that closely mirror the training trends, demonstrating good generalization and minimal overfitting. Validation losses (box, classification, and DFL) decrease progressively, while performance metrics like mAP@0.5 and mAP@0.5:0.95 increase steadily—reaching around 0.93 and 0.79, respectively. These high mAP scores indicate that the model is not only correctly identifying objects but also predicting their bounding boxes with strong accuracy. Overall, the results suggest a well-trained and reliable YOLOv11 model suitable for deployment in real-world detection tasks.

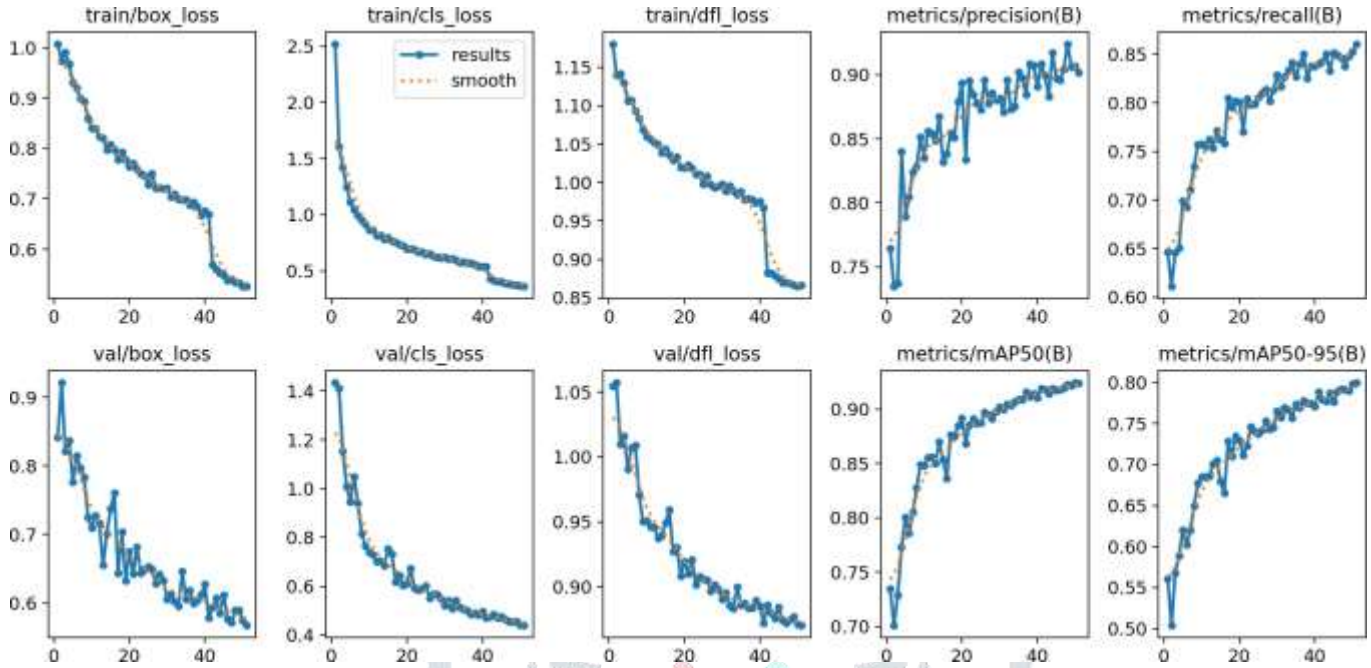

Fig. 2: Peformance of YOLOv11

TABLE III: YOLO 11 Architecture on Pomegranate Growth Stages

Metric/Feature	**YOLOv11**
Training Losses (Mean)	Detailed insights available, generally decreasing trends across epochs.
Validation Losses (Mean)	Reflects consistent generalization trends with decreasing validation losses.
Evaluation Metrics	Consistently high metrics mAP50: 0.995 mAP50-95: 0.995
F1-Confidence Curve	Peak F1: 0.99 Threshold: 0.76
Precision-Confidence Curve	Precision: 1.00 Threshold: 0.942
PR Curve (Precision-Recall)	mAP: 0.995 High precision and recall maintained consistently
Recall-Confidence Curve	Recall: 1.00 Gradual decline at higher confidence
Epoch Time (Mean)	Detailed, showing computational efficiency and adjustments.

VII. Conclusion and Future Scope

The proposed YOLOv11-based model effectively detects and classifies the different growth stages of pomegranates with high accuracy and robust performance across key metrics such as precision, recall, and mAP. By automating the identification process, this system overcomes the limitations of manual

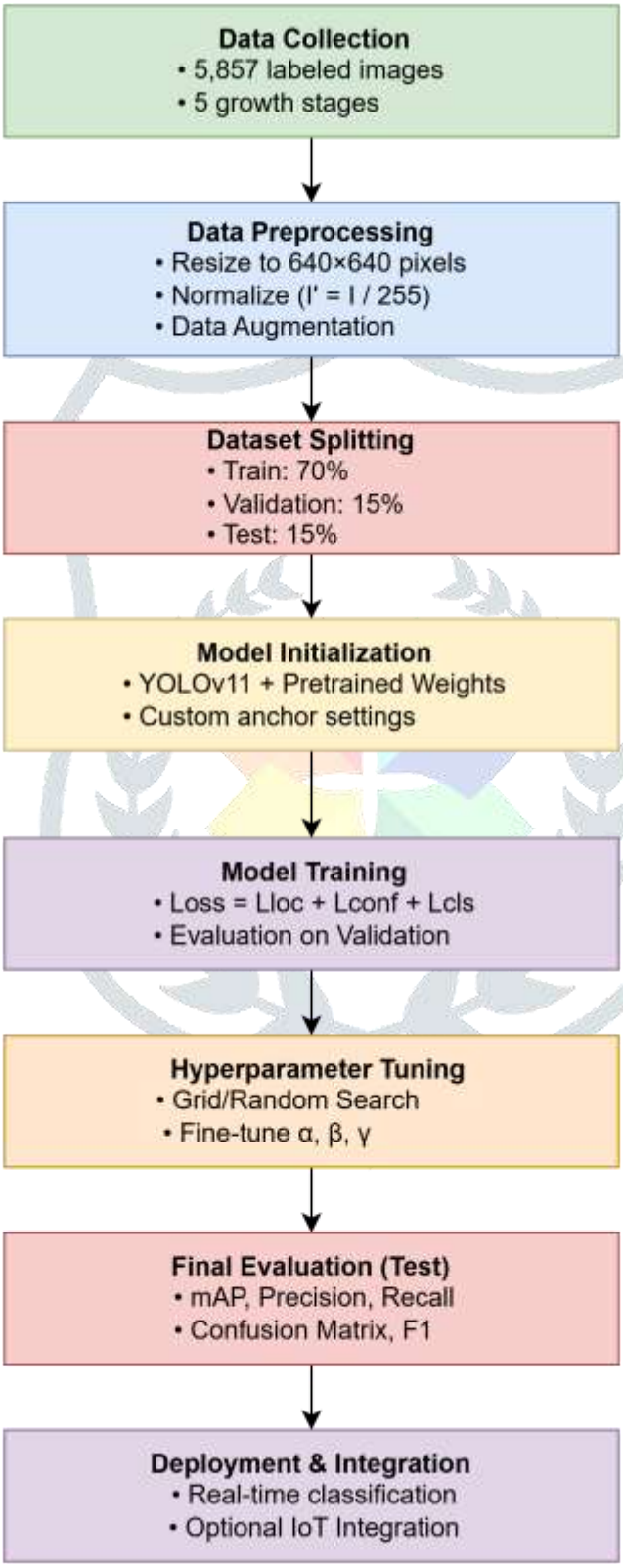

Fig. 3: Proposed Methodology

monitoring, enabling timely and precise crop management decisions that can improve yield quality and resource efficiency. The model's ability to generalize well to unseen data, aided by data augmentation and hyperparameter tuning, demonstrates its potential as a practical tool for precision agriculture in real-world farming scenarios.

For future work, expanding the dataset with more diverse images and integrating multimodal data such as hyperspectral or thermal imaging could enhance detection accuracy and resilience under varying environmental conditions. Additionally, optimizing the model for deployment on edge devices and drones would facilitate real-time, on-site monitoring. Further development could also include integration with automated farm management systems and continuous learning capabilities to adapt to seasonal changes. Field trials and feedback from end-users will be essential to refine the system and fully realize its benefits for sustainable pomegranate cultivation.

REFERENCES

[1] Akhilesh Sharma, Vipan Kumar, and Louis Longchamps. Comparative performance of yolov8, yolov9, yolov10, yolov11 and faster r-cnn models for detection of multiple weed species. *Smart Agricultural Technology*, 9:100648, 2024.

[2] Gustavo PCP da Luz, Gabriel Massuyoshi Sato, Luis Fernando Gomez Gonzalez, and Juliana Freitag Borin. Smart parking with pixel-wise roi selection for vehicle detection using yolov8, yolov9, yolov10, and yolov11. *arXiv preprint arXiv:2412.01983*, 2024.

[3] Alok Ranjan Sahoo, Satya Sangram Sahoo, and Pavan Chakraborty. Polyp detection in colonoscopy images using yolov11. *arXiv preprint arXiv:2501.09051*, 2025.

[4] Yaxian Lei, Xiaochong Tong, Chunping Qiu, Haoshuai Song, Congzhou Guo, and He Li. Spatial-aware remote sensing image generation from spatial relationship descriptions. *IEEE Geoscience and Remote Sensing Letters*, 2025.

[5] Yong-Hwan Lee and Heung-Jun Kim. Comparative analysis of yolo series (from v1 to v11) and their application in computer vision. *Journal of the Semiconductor & Display Technology*, 23(4):190–198, 2024.

[6] Liyang Xiao, Yanni Yang, Zhe Chen, Gao Yue, Prasant Mohapatra, and Pengfei Hu. Crfusion: Fine-grained object identification using rf-image modality fusion. *IEEE Transactions on Mobile Computing*, 2025.

[7] Hatim Derrouz, Hamza Alami, and Reda Rabie. Mlpr: Yolov3 for real-time license plate recognition in moroccan video streams. *IEEE Access*, 2025.

[8] Jinfan Wei, Lingyun Ni, Lan Luo, Mengchao Chen, Minghui You, Yu Sun, and Tianli Hu. Gfs-yolo11: A maturity detection model for multi-variety tomato. *Agronomy*, 14(11):2644, 2024.

[9] Ming Him Lui, Haixu Liu, Zhuochen Tang, Hang Yuan, David Williams, Dongjin Lee, KC Wong, and Zihao Wang. An adaptive yolo11 framework for the localisation, tracking, and imaging of small aerial targets using a pan–tilt–zoom camera network. *Eng*, 5(4):3488–3516, 2024.

[10] Ali Ghahremani, Scott D Adams, Michael Norton, Sui Yang Khoo, and Abbas Z Kouzani. Detecting defects in solar panels using the yolo v10 and v11 algorithms. *Electronics*, 14(2):344, 2025.