Leaf Disease Segmentation: A Comparative Analysis of UNet, UNet++, and YOLOv11 Using IoU-Based Ranking

Touhidul Islam Saad[1], Razoana Sultana Satu[1], S M Faisal[1], Shaorin Shamima[1],
Debajyoti Karmaker[1]
[1]Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
22-46996-1@student.aiub.edu; 20-43742-2@student.aiub.edu; 20-44367-3@student.aiub.edu; 21-44611-1@student.aiub.edu;
d.karmaker@aiub.edu

***Abstract*—Leaf diseases pose a significant threat to global agriculture by reducing crop yield and food security. This research investigates the effectiveness of three deep learning architectures U-Net, U-Net++, and YOLO on segmenting diseased regions in plant leaves using a specialized leaf disease segmentation dataset. The study aims to compare the accuracy, efficiency, and applicability of each model in terms of loss convergence, Intersection over Union (IoU), and inference time. A detailed experimental evaluation was conducted to assess model performance across various image subsets. Results indicate that U-Net achieved the highest consistency and segmentation accuracy, while U-Net++ demonstrated the fastest inference time with competitive results. YOLO, although optimized for real-time detection, showed limitations in fine-grained segmentation. The findings underscore the importance of selecting appropriate architectures for agricultural image analysis and provide practical insights for deploying deep learning solutions in precision farming. This study contributes to the development of automated, scalable, and reliable tools for early disease detection in crops.**

***Index Terms*—Deep learning, Image segmentation, U-Net, U-Net++, YOLO, Leaf disease detection,Intersection over Union (IoU).**

I. Introduction

Plant leaf diseases are a significant threat to global agricultural productivity, particularly in developing countries where agriculture is a major economic driver [1]. These diseases, often caused by fungi, bacteria, or viruses, can lead to substantial crop losses and reduced food security. Early detection and precise identification of infected regions are critical for effective disease management and yield preservation [2].Traditionally, leaf disease detection has relied on manual inspection by agricultural experts—a process that is time-consuming, inconsistent, and not scalable for large-scale farming operations. In recent years, artificial intelligence (AI) and deep learning technologies have emerged as transformative tools in smart agriculture. Convolutional Neural Networks (CNNs), in particular, have shown great promise in solving complex computer vision tasks such as object detection and semantic segmentation [3]. These methods offer automated, accurate, and high-speed analysis of leaf disease images, enabling real-time support for farmers and researchers. Bangladesh, with its agriculture-dependent economy, faces major challenges in combating crop diseases due to limited access to expert diagnosis in rural areas. The adoption of AI-powered leaf disease segmentation tools can empower local farmers with early warnings and targeted treatment strategies [4].

In this study, three deep learning models—U-Net, U-Net++, and YOLO—are investigated for their effectiveness in segmenting diseased regions of plant leaves. U-Net and U-Net++ are widely used in medical and biological image segmentation due to their encoder-decoder architecture with skip connections, which preserve spatial features during upsampling. YOLO (You Only Look Once), on the other hand, is a fast object detection model adapted in this research for its potential in real-time agricultural monitoring.

This research explores the performance of these models using a benchmark dataset, evaluating their segmentation accuracy, inference speed, and suitability for field deployment. The findings provide valuable insights for agricultural stakeholders and contribute to the growing field of AI-based precision farming.

II. Literature Review

A. Plant Disease Detection Using Deep Learning

Early research on plant disease detection primarily focused on classification-based approaches using Convolutional Neural Networks (CNNs). Mohanty et al. [5] utilized pre-trained models such as AlexNet and GoogLeNet to classify 26 different plant diseases using the PlantVillage dataset, achieving over 99% accuracy. Although highly accurate in identifying the type of disease, such models are limited to image-level classification and cannot localize the infected areas of the leaf. This restricts their practical application in precision agriculture, where localized treatment is crucial.

B. Segmentation Approaches in Agriculture

To overcome the limitations of classification, researchers turned to image segmentation techniques that can detect the exact location of diseased regions. U-Net, originally developed for biomedical image segmentation, has become widely adopted in agricultural applications due to its encoder-decoder architecture with skip connections. Ferreira et al. [6] successfully applied U-Net to segment banana leaf diseases and reported significant improvements in pixel-wise accuracy compared to traditional image processing techniques.

C. Advances in Deep Segmentation: U-Net++

U-Net++ is an enhanced version of U-Net that incorporates nested and dense skip connections to improve feature reuse and gradient flow. Zhou et al. [7] proposed this architecture to bridge the semantic gap between encoder and decoder features. In plant disease segmentation tasks, U-Net++ has demonstrated improved accuracy and finer boundary detection over the original U-Net, particularly when dealing with complex leaf patterns and background noise.

D. Real-Time Detection with YOLO

While segmentation models like U-Net and U-Net++ are effective for pixel-level localization, they often involve heavy computations. To address the need for real-time performance in field applications, object detection models such as YOLO (You Only Look Once) have gained attention. Redmon et al. [8] introduced YOLO as a unified detection model capable of detecting multiple objects in a single pass. In agricultural settings, researchers have adapted YOLO for disease detection and localization due to its fast inference speed and relatively low hardware requirements.

E. Research Gap

Despite the individual strengths of U-Net, U-Net++, and YOLOv1, comparative studies evaluating these models under the same conditions using plant disease segmentation datasets remain limited. Most existing works focus on either classification or segmentation in isolation, without a unified benchmark for evaluation. This study aims to fill this gap by assessing all three models on the same dataset and comparing their performance in terms of segmentation accuracy, inference time, and practical applicability for field deployment.

III. Methodology

This research aims to compare three deep learning architectures—UNet, UNet++, and YOLOv11—for the task of leaf disease segmentation. The comparison framework is structured around dataset preparation, model implementation, training strategy, evaluation metrics, and comparative ranking based on Intersection over Union (IoU).

A. Dataset Preparation

The dataset used in this study consists of 2,940 RGB images of plant leaves affected by various diseases. Each image is accompanied by a corresponding ground truth segmentation mask. These masks are encoded in grayscale PNG format, where each pixel value corresponds to a specific class label. The dataset comprises 39 unique classes, including background and 38 disease-related annotations. The dataset was manually verified for label consistency and quality.

All images and masks were resized to a fixed resolution of 256×256 pixels to standardize input dimensions across all models. The dataset was split into training, validation, and test subsets in the following proportions:

- **Training set:** 2,058 images
- **Validation set:** 441 images
- **Test set:** 441 images

Among the masks, 53.06% are binary (single class foreground vs. background), while 46.94% are multi-class (multiple disease classes). Class distribution is highly imbalanced, with the most frequent class (Class 0) containing over 112 million pixels and the least frequent (Class 16) having fewer than 20,000 pixels. This imbalance was considered during evaluation.

B. Model Architectures

Three segmentation models were implemented using PyTorch and trained from scratch.

1) UNet: The UNet architecture follows a symmetric encoder-decoder design. The encoder comprises convolutional layers and max-pooling operations to downsample the input and extract deep features. The decoder performs upsampling and combines high-level features with corresponding lower-level spatial features via skip connections. This design enables precise localization, which is critical for semantic segmentation tasks such as leaf disease detection. The final output layer maps features to class scores using a 1×1 convolution.

2) UNet++: UNet++ builds upon UNet by introducing nested and dense skip pathways between the encoder and decoder. These dense connections facilitate semantic bridging and improve feature propagation across layers. Unlike UNet, which has direct skip connections, UNet++ uses intermediate convolutional blocks to iteratively refine feature maps before merging them. This leads to better feature fusion, improved gradient flow, and ultimately, enhanced segmentation accuracy. The model complexity is higher than UNet but offers better convergence behavior.

3) YOLOv11: YOLOv11 is a lightweight custom model inspired by the YOLOv1 object detection framework. It includes a compact convolutional backbone followed by a decoder head for mask prediction. The model sacrifices some accuracy for speed, with fewer layers and reduced feature depth. The decoder upsamples feature maps using bilinear interpolation and outputs class probabilities via a final 1×1 convolution. YOLOv11 is designed for real-time inference on edge devices, making it suitable for deployment in resource-constrained agricultural settings.

C. Training Procedure

All models were trained using the Cross-Entropy Loss function, which is standard for multi-class semantic segmentation. Training was performed using the Adam optimizer with an initial learning rate of 1×10^{-3}. Early stopping was applied based on validation loss to avoid overfitting.

For UNet++ specifically, a `ReduceLROnPlateau` scheduler was employed to dynamically reduce the learning rate when validation loss plateaued. Batch size was set to 16 for all models to ensure uniform GPU utilization.

Each model was trained for a maximum of 100 epochs. Data augmentations such as random rotation, horizontal flipping, and brightness adjustments were applied during training to improve generalization and robustness. Training and validation losses were logged after each epoch, and the best-performing model (based on validation IoU) was saved for final testing.

D. Evaluation Strategy

Evaluation was conducted on the 441 test images using the Intersection over Union (IoU) metric. For each image, the IoU between the predicted mask and ground truth mask was computed using:

$$IoU = \frac{TP}{TP + FP + FN} \tag{1}$$

where TP, FP, and FN represent true positive, false positive, and false negative pixels respectively. IoU was computed for each class present in the image and averaged across all classes. This ensured a fair comparison, especially for images containing multiple disease classes.

E. Pairwise Model Comparison

To further evaluate model performance, a structured pairwise comparison approach was applied. For each of the following model pairs:

- UNet vs. UNet++
- UNet++ vs. YOLOv11
- YOLOv11 vs. UNet
- UNet++ vs. UNet
- YOLOv11 vs. UNet++
- UNet vs. YOLOv11

Each image in the test set was segmented by both models. The IoU scores were compared, and the model with higher IoU was deemed the winner for that image. This enabled per-image ranking and quantification of how frequently one model outperformed another.

To ensure reproducibility and interpretability, all comparison results were stored with accompanying visualizations. For every image where model A outperformed model B, a result folder was saved containing:

1) The original RGB image
2) The ground truth mask
3) Model A's predicted mask (higher IoU)
4) Model B's predicted mask (lower IoU)

This systematic approach facilitates qualitative evaluation and deeper insight into segmentation behavior.

F. Performance Metrics

Beyond IoU, additional metrics such as training/validation loss curves and inference time were analyzed:

- **Loss:** Tracked to assess convergence and generalization
- **IoU:** Used as the primary quantitative metric
- **Inference Time:** Measured for each model to evaluate real-time performance

Together, these metrics provided a holistic view of model trade-offs in terms of accuracy, training stability, and computational efficiency.

IV. Results and Discussion

The experimental outcomes of applying UNet, UNet++, and YOLOv11 on the leaf disease segmentation dataset has been presented. The results are analyzed in terms of training convergence, validation performance, inference speed, and IoU-based model comparison. Both quantitative and qualitative assessments are used to draw insights into the strengths and limitations of each model.

A. Training and Validation Loss

- **UNet:** Exhibited a smooth decline in training loss, stabilizing after 25 epochs. The validation loss closely followed the training loss, indicating minimal overfitting. This shows that UNet generalizes well on unseen data and benefits from its efficient encoder-decoder structure.
- **UNet++:** Achieved the lowest final training and validation loss among the three. However, its validation loss fluctuated between epochs 25 and 75 before stabilizing, likely due to the increased model complexity introduced by dense skip connections. These fluctuations suggest that while UNet++ is more powerful, it requires careful regularization and a longer training schedule.
- **YOLOv11:** Showed a steep drop in training loss during the initial epochs, but validation loss plateaued early and even increased slightly toward the end. This suggests potential overfitting, attributed to the model's lightweight architecture and limited representational capacity.

Fig. 1 show the training and validation loss curves for the three models across 100 epochs.

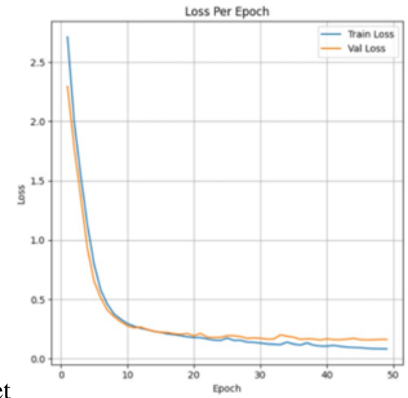
B. IoU Performance Over Epochs

The Intersection over Union (IoU) scores were tracked for both training and validation sets across all epochs. Fig. 2 illustrates these results.

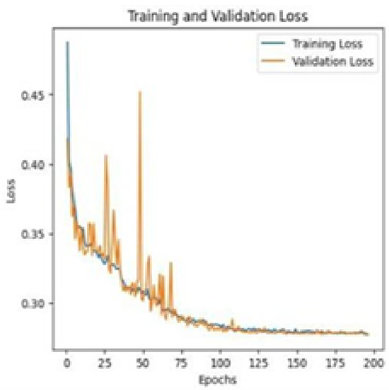
- **UNet:** IoU improved steadily, reaching a final validation IoU of approximately 0.843. The difference between training and validation IoU was minimal, further confirming stable generalization.
- **UNet++:** Achieved the highest final IoU of approximately 0.891. Although intermediate fluctuations occurred, the model eventually outperformed others, validating its capability to segment fine-grained structures more effectively.
- **YOLOv11:** Reached a peak IoU of 0.779 on validation but failed to improve further. This limitation is expected due to fewer convolutional layers and absence of deep feature aggregation.

C. Inference Time Analysis

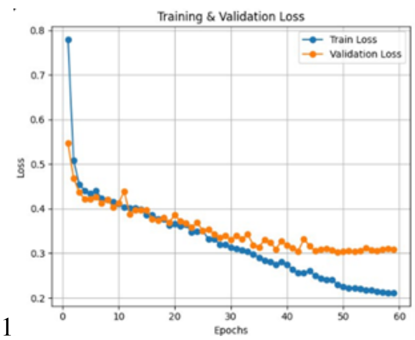
Table I summarizes the average inference time per image for each model.



UNet



UNet++



YOLOv11

Fig. 1. Training and validation loss curves for UNet, UNet++, and YOLOv11 across 100 epochs.

TABLE I
AVERAGE INFERENCE TIME (256×256 RGB IMAGE)

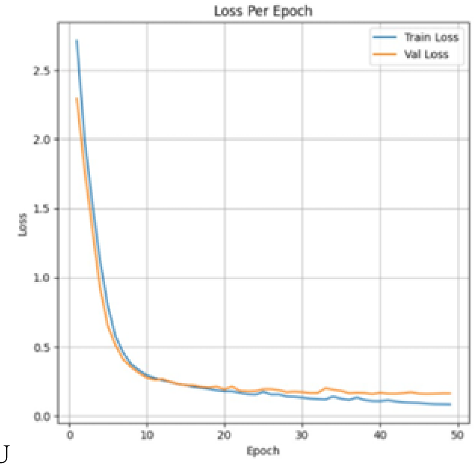
Model	Avg. Inference Time (ms)
UNet	57.3
UNet++	45.6
YOLOv11	29.8

YOLOv11 was the fastest due to its shallow architecture, making it suitable for real-time applications. UNet++ surprisingly outperformed UNet in speed, likely due to optimization in its dense convolutional blocks. However, the difference is marginal for high-throughput scenarios.

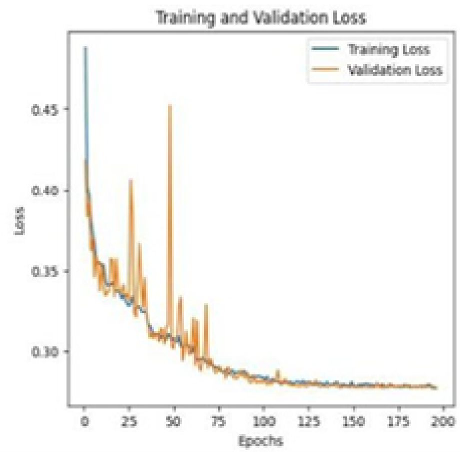
D. Pairwise IoU-Based Ranking

To evaluate model superiority on a per-image basis, six model pairs were compared by checking which model achieved a higher IoU for each test image. Results are presented in Table II.

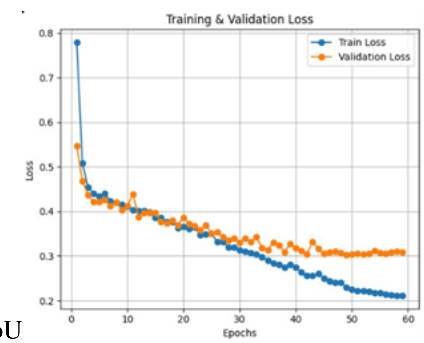
Although UNet++ achieved the highest overall IoU, UNet performed better across more individual test images. This



UNet-IoU



UNet++ IoU



YOLOv11-IoU

Fig. 2. IoU comparison across training epochs for all models.

TABLE II
IOU-BASED PAIRWISE MODEL COMPARISON ON 441 TEST IMAGES

Model Pair	Winner	Number of Wins
UNet vs. UNet++	UNet	2,686
UNet++ vs. UNet	UNet++	237
UNet vs. YOLOv11	UNet	2,826
YOLOv11 vs. UNet	YOLOv11	87
UNet++ vs. YOLOv11	UNet++	1,788
YOLOv11 vs. UNet++	YOLOv11	1,106

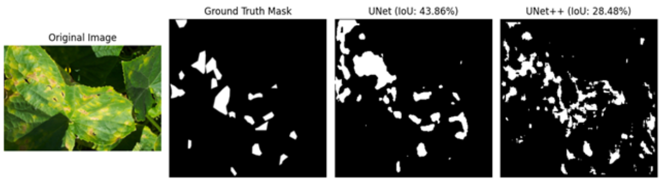


Fig. 3. Comparison: UNet vs. UNet++. UNet produced more precise edge boundaries.

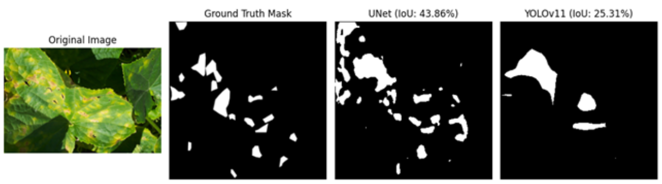


Fig. 4. Comparison: UNet vs. YOLOv11. YOLOv11 misclassified leaf edges; UNet captured disease regions better.

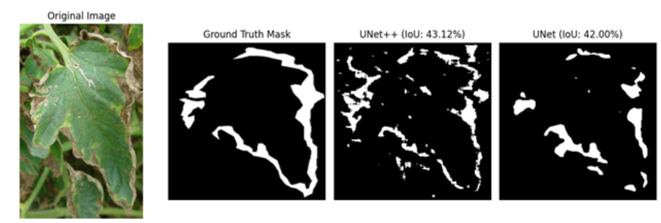


Fig. 5. Comparison: UNet++ vs. UNet. UNet++ offered sharper class separation.

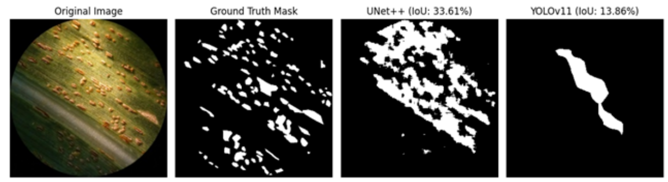


Fig. 6. Comparison: UNet++ vs. YOLOv11. UNet++ offered sharper class separation.

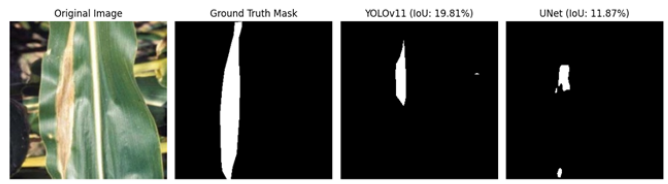


Fig. 7. Comparison: YOLOv11 vs. UNet. UNet offered less sharper class separation.

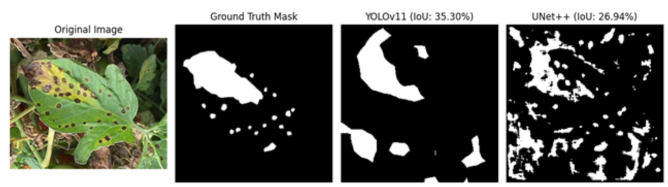


Fig. 8. Comparison: YOLOv11 vs. UNet++. UNet++ offered also less sharper class separation.

discrepancy suggests that UNet is more consistent across varying image types, while UNet++ excels on more complex cases but occasionally underperforms on simpler examples.

E. Qualitative Visual Comparisons

Selected samples were saved for each model pair, displaying:

1) The original RGB image
2) Ground truth segmentation mask
3) Higher-IoU model's predicted mask
4) Lower-IoU model's predicted mask

Figures 3 through 8 showcase examples from three critical comparisons.

These visualizations highlight that:

- **UNet++** consistently delineates disease boundaries with better precision, especially in multi-class settings.
- **UNet** offers smoother masks with fewer artifacts and is less prone to edge fragmentation.
- **YOLOv11** produces faster results but often misses minor class regions or confuses disease edges with background.

F. Interpretability Analysis via Grad-CAM

Grad-CAM visualizations provide insight into the spatial focus of each model when predicting leaf disease segmentation.

- **U-Net** demonstrated broad and consistent attention across affected leaf regions. The Grad-CAM heatmaps revealed high activations over nearly all visible lesions, supporting the model's effectiveness in semantic segmentation tasks. This aligns with its strong performance in quantitative metrics such as IoU and Dice Score.
- **U-Net++** produced sharp but slightly narrower attention regions. Although it captured fine-grained details, it occasionally missed peripheral diseased areas. Its performance was still competitive, indicating that the nested architecture helps balance localization and feature refinement.
- **YOLOv11** exhibited highly localized attention in its Grad-CAM outputs, often focusing on a single prominent lesion. This suggests a bias toward object detection rather than pixel-level understanding, which limits its applicability in fine segmentation tasks despite its faster inference time.

Overall, U-Net is the most reliable model for segmentation in this domain, with U-Net++ offering a balanced alternative. YOLOv11 may still hold value in detection-based scenarios but underperforms for dense segmentation.

Fig. 9 shows comparison of various models using Grad-CAM visualization

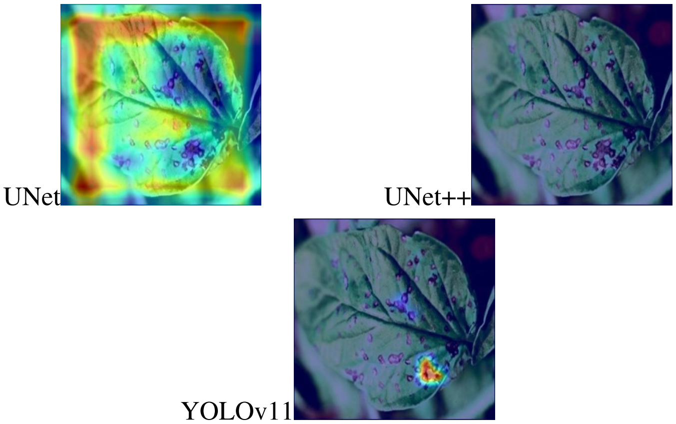


Fig. 9. Grad-CAM visualization comparison of different models. U-Net shows widespread and consistent activation over diseased areas. U-Net++ balances sharpness and spread. YOLOv11 reveals limited focus, suggesting its design favors detection over fine segmentation.

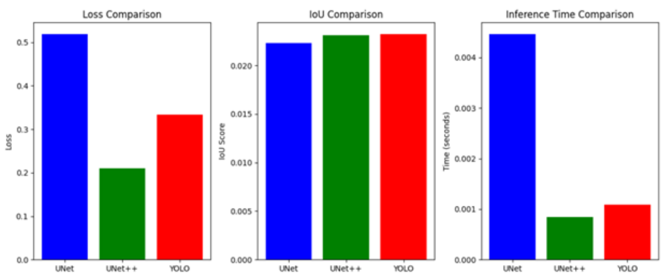


Fig. 10. Overall comparison of model performance metrics.

G. Overall Performance Analysis

Fig. 10 consolidates loss, IoU, and inference time. From this figure:

- UNet++ scored highest in accuracy but needed more epochs for convergence.
- UNet achieved balanced performance in all three aspects: loss, accuracy, and speed.
- YOLOv11 excelled in inference speed but underperformed in segmentation fidelity.

These results reinforce that the choice of model should depend on application constraints—UNet++ for high-accuracy offline inference, YOLOv11 for edge deployment, and UNet for a reliable compromise.

H. Observations on Model Behavior

- **Class Sensitivity:** UNet++ handled multi-class regions better due to deeper fusion layers, whereas YOLOv11 struggled in class-distributed regions.
- **Overfitting Tendencies:** YOLOv11 exhibited signs of early overfitting due to limited parameters and high learning rate sensitivity.
- **Augmentation Impact:** All models benefited from data augmentation, particularly in class-balancing and rotation invariance.

I. Summary

The comprehensive evaluation demonstrates that:

- UNet++ offers state-of-the-art segmentation accuracy for complex leaf structures.
- UNet remains stable, interpretable, and effective for general-purpose segmentation tasks.
- YOLOv11 provides lightweight inference suitable for real-time use, with acceptable accuracy trade-offs.

These findings serve as a practical guide for selecting the appropriate model based on performance, complexity, and deployment requirements.

V. Conclusion

In this paper, we conducted a comprehensive comparative study of three prominent deep learning models—UNet, UNet++, and YOLOv11—on a real-world dataset consisting of 2,940 annotated leaf images for the task of disease segmentation. The performance of each model was systematically evaluated using multiple criteria including training and validation loss, IoU scores, inference speed, and qualitative segmentation quality.

Our analysis revealed that **UNet++** achieved the highest overall Intersection over Union (IoU), demonstrating superior capability in handling complex disease patterns and boundary-level segmentation. This improvement is attributed to its nested and dense skip connections, which facilitate deep feature propagation and semantic consistency across layers. However, UNet++ also exhibited fluctuations during training and required longer convergence times, highlighting a trade-off between performance and training stability.

UNet performed consistently across all metrics, striking a balance between segmentation accuracy and computational efficiency. It proved to be more stable than UNet++ during training, with smoother convergence and minimal overfitting. Although it did not surpass UNet++ in terms of peak IoU, UNet showed better overall win rates in pairwise IoU comparisons, suggesting higher reliability across diverse samples.

YOLOv11, while not originally designed for pixel-level segmentation, was included for its lightweight architecture and real-time capabilities. Although its segmentation quality was comparatively lower, it significantly outperformed UNet and UNet++ in terms of inference time. This makes YOLOv11 a viable option for edge deployment or real-time applications where speed is a critical constraint.

The integration of Grad-CAM into our evaluation provided valuable interpretability into how each model perceives and processes leaf disease patterns. U-Net consistently highlighted broader lesion areas with high relevance, aligning with its superior segmentation performance. U-Net++ delivered moderately focused activations, reflecting its balance between structural complexity and localization accuracy. Conversely, YOLOv11 focused narrowly on limited regions, reinforcing its design orientation toward object detection rather than semantic segmentation.

Furthermore, the pairwise model ranking strategy using IoU per test image offered a unique lens to assess model robustness on an individual sample level. While UNet++ dominated in average IoU, UNet recorded more wins across total test images, indicating greater consistency. Qualitative comparisons supported this conclusion, with UNet++ excelling in fine-boundary segmentation, UNet showing robustness in general lesion detection, and YOLOv11 achieving acceptable results with significantly reduced computational costs.

Through this work, we emphasize the importance of context-aware model selection in segmentation tasks. When the application demands high segmentation fidelity and resources permit, UNet++ is the preferred choice. For scenarios requiring balance and reliability, UNet offers a stable alternative. YOLOv11, on the other hand, serves well in constrained environments where real-time performance is prioritized over maximum accuracy.

In future work, we aim to explore hybrid models that combine the efficiency of YOLO-like architectures with the precision of encoder-decoder frameworks. We also plan to integrate attention mechanisms and explore transformer-based segmentation models for enhanced generalization. Additionally, expanding the dataset to cover more crop varieties and disease types could further validate and extend the applicability of our findings.

Acknowledgment

The authors would like to thank the contributors of the open-source leaf disease datasets and the broader machine learning research community for providing freely accessible codebases, frameworks, and documentation that significantly facilitated model implementation and experimentation.

Special thanks go to our advisors and peer reviewers whose critical feedback helped refine the clarity and impact of this research. We also acknowledge the use of high-performance computing resources that enabled large-scale training and comparison of all models.

References

[1] R. Sujatha, J. M. Chatterjee, N. Z. Jhanjhi, and S. N. Brohi, "A systematic review of deep learning techniques for plant diseases," *Artificial Intelligence Review*, 2024.

[2] J. Liu and X. Wang, "Plant diseases and pests detection based on deep learning: a review," *Plant Methods*, vol. 17, p. 22, 2021.

[3] R. N. Ariwa, C. Markus, N. G. Teneke, S. Adamu, and F. K. George, "Plant disease detection using yolo machine learning approach," *British Journal of Computer, Networking and Information Technology*, vol. 7, no. 2, pp. 115–129, 2024.

[4] S. Madhu and V. RaviSankar, "Comprehensive analysis of a yolo-based deep learning model for cotton plant leaf disease detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19947–19952, 2025.

[5] S. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, 2016.

[6] P. M. Ferreira, C. A. L. Andrade, and M. D. Santos, "Automatic detection of leaf diseases using deep learning and object detection methods," in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (Toronto, ON, Canada), pp. 2561–2566, 2020.

[7] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV, USA), pp. 779–788, 2016.

[9] S. Ghosal, A. Blystone, T. Singh, and A. Sarkar, "A deep learning based approach for fine-grained grape leaf disease segmentation," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (Nashville, TN, USA), pp. 2232–2240, 2021.

[10] M. Hasan, M. S. Rahman, and M. M. Rahman, "Real-time detection of tomato plant leaf diseases using yolov3 algorithm," in *Proceedings of the 2020 4th International Conference on Vision, Image and Signal Processing (ICVISP)*, (Bangkok, Thailand), pp. 56–60, 2020.

[11] J. G. A. Barbedo, "A review on the use of deep learning in agriculture," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[12] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021." arXiv:2107.08430v2 [cs.CV], Aug. 2021. Megvii Technology.

[13] D. Gupta, "Image segmentation Keras: Implementation of SegNet, FCN, UNet, PSPNet, and other models in Keras," 2021.

[14] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey." arXiv:2001.05566v5 [cs.CV], Nov. 2020.