Vision-Based YOLOv5 Object Detection and Prediction of Berry Blotch Disease Severity Estimation in Coffee Plantation

S. Raveena[1], Midhunchakkaravarthy[2], Eugenio Vocaturo[3]

[1] Department of Computer Science and Engineering, Lincoln University College, Petaling Jaya 47301, Malaysia;

[2] Faculty of Computer Science & Multimedia, Lincoln University College, Petaling Jaya 47301, Malaysia;

[3] DIMES - Department of Computer Engineering, Electronic Modeling, and Systems Engineering, University of Calabria, Rende, Italy

Email ID: midhun@lincoln.edu.my; raveenaselvanarayanan@gmail.com

---

Abstract: Berry blotch disease adversely impacts coffee agriculture, creating significant economic losses from dwindling both yield and quality. The present investigation presents a novel vision-based methodology employing YOLOv5 object detection to identify and segment diseased coffee berries, subsequently assessing disease severity through a hybrid deep learning framework that integrates Convolutional Neural Networks (CNN) and Vision Transformer-CNN (ViT-CNN). The YOLOv5 model demonstrates effective identification of diseased berries, using improvements in speed and accuracy compared to YOLOv3. Post-detection, the hybrid ViT-CNN model accurately diagnoses the disease's severity levels by integrating spatial and contextual data. The suggested approach has an accuracy of 95.7%, outperforming current techniques in detection and classification tasks. This comprehensive system offers a scalable and effective solution for real-time disease surveillance and management in coffee plantations, allowing prompt action to reduce crop losses and promote sustainable agricultural practices.

Keywords: Berry blotch disease, YOLOv5, ViT-CNN, Object detection, Severity estimation, Disease monitoring, Real-time detection.

Introduction

Berry blotch disease reduces coffee bean production by inducing early berry fall. Bean Quality Decline Infected beans may not meet quality standards for higher prices. Farmers must buy and apply fungicides and utilize cultural methods to control disease. Coffee farmers may face financial instability due to low yields and low-quality beans. Coffee farming is a major cash source for smallholder farmers, and disease outbreaks threaten their food security [1]. Symptoms include Unusual slightly deep necrotic patches of dark brown coloration on the exposed surface of green berries. Large necrotic lesions might cover the berries. A purple halo may surround necrotic regions. Berry blotch, a common fungal disease, reduces coffee plantation yield and bean quality. Successful disease treatment requires accurate and timely sickness severity assessment. Traditional sickness severity assessments use qualified professionals' eye exams, which may be laborious, subjective, and error-prone. Due to these limits, more objective and efficient methods for measuring sickness severity have been developed [2]. Recent improvements in image processing and machine learning have created new opportunities for the development of automated systems for illness severity assessment. These systems employ computer vision methods to examine images of coffee cherries and detect and measure the degree of infection. YOLOv5 is a cutting-edge object identification algorithm recognized for its rapidity and precision. The model accepts an image as input. Feature extraction image undergoes processing via many layers of a convolutional neural network (CNN). These layers extract significant elements such as edges, forms, and textures. The prediction grid image is segmented into a grid of cells. Each cell forecasts the existence of items included inside it, together with their bounding box coordinates and class probability. The object detection model evaluates the predictions from each cell and recognizes items inside the picture. It identifies the object's category (e.g., leaves, flowers, coffee cherries) and delineates bounding boxes around them. The final result comprises an image including bounding boxes around the identified items, along with their class labels and confidence ratings [3]. A Coffee Berry Blotch Disease Prediction System combines computer vision and machine learning to precisely and effectively identify and evaluate the severity of this fungal affliction in coffee plants. The approach often entails acquiring high-resolution photographs of coffee cherries via the use of cameras or drones. The photos are further preprocessed to boost quality and subjected to data augmentation to bolster model resilience [4]. Feature extraction techniques, including manual methods (e.g., color histograms, texture descriptors) and deep learning approaches (e.g., convolutional neural networks such as YOLOv5, EfficientNet, and ResNet), are used to discern critical attributes of the illness. Machine learning methods, such as classifiers (e.g., SVM, Random Forest) and regression models (e.g., Linear Regression, Support Vector Regression), are developed to categorize pictures as healthy or infected and to assess the degree of infection. The trained model is then implemented on an appropriate platform (e.g., cloud server, edge device) and linked with an intuitive interface for farmers to contribute photos and get forecasts. This system has several benefits, such as early illness identification, enhanced disease management techniques, heightened production, and less dependence on manual inspections. Nonetheless, obstacles like data acquisition, image variation, and computing resource demands must be resolved for effective deployment and broad recognition [5].

Related work

West African taro farmers face food security and economic instability due to Taro Leaf Blight (TLB). For early TLB identification in taro plants, the YOLOv8 deep learning model is used for object recognition. Taro leaves at different infection stages from Nigerian and Ghanaian farms were photographed for the research. Enhancing the YOLOv8 model for this dataset yielded 85.7% mean Average Precision (mAP) across all classes, beating conventional plant disease detection methods that typically reach 70-75% on similar datasets. The 15-20% increase in early detection accuracy allows faster treatments. Farmers received real-time disease monitoring and testing using an Android app. The method proved effective and usable in field testing, making it a viable early illness intervention tool. Deep learning and mobile technologies may solve agricultural problems and boost regional food security [6]. Safeguarding plants from illnesses requires identifying symptoms and finding safe, effective treatments. Seasonal holistic strategies may prevent plant resistance and unnecessary duties. Technological advances have made leaf image disease diagnosis fast and automated using machine learning. This research used computer vision neural networks EfficientNet and YOLO to create a prediction model. A program that helps farmers diagnose cotton diseases and apply pesticides is based on this idea. Model output must be monitored in the physical world, where information comes from many sources. Monitoring affects model reaction to actual device inputs, according to the study. A new convolutional neural network classifies and recognizes cotton leaves using EfficientNet and YOLO architectures. EfficientNetB4 classified healthy and powdery mildew leaves with 100% accuracy, whereas YOLO v4 had 96% precision, recall, mAP@0.5, 99.2%, and 0.70 [7]. The automatic identification of downy mildew in grapes, apple scab in apples, and Alternaria leaf blight in carrots, using a deep convolutional neural network (CNN) on RGB color photographs. Outputs from the CNN were used as input for a Decision Support System (DSS) to accurately identify and measure the illness, enabling the recommendation of suitable and timely application of plant protection measures. Concentrated on installing a smart camera including integrated deep-learning processing in real-world situations [8]. Bananas, a vital fruit, are afflicted by diseases. In the absence of early intervention, East Africa's most significant banana diseases, Fusarium Wilt and Black Sigatoka, may devastate 30% to 100% of the harvests. Efficient control of banana diseases requires prompt identification to mitigate production and economic losses. This initiative was motivated by recent advances in deep learning for plant disease identification. They used a U-Net semantic segmentation deep learning network for the early identification and segmentation of Fusarium Wilt and Black Sigatoka banana diseases. These two illnesses impacted 18,240 photos of banana leaves and stalks used for training this model. Cell phones were used to gather and classify agricultural photographs alongside professionals. The U-Net model has a Dice Coefficient of 96.45% and an Intersection over Union (IoU) of 93.23%. The model precisely delineated banana leaves and stems affected by Fusarium Wilt and Black Sigatoka [9]. The soybean plant leaf disease dataset from actual fields mitigates the limitations of existing collections. Experimental results evaluate the proposed model against 10 leading methodologies across five metrics: accuracy, precision, recall, and F1-score. The model's efficacy is validated using four public datasets: Embrapa, Plant Village, AI2018, and PlantDoc. The proposed model surpassed 10 advanced models, achieving accuracies of 98.00%, 97.00%, 76.00%, and 92.00% on the Plant Village, AI2018, PlantDoc, and Embrapa datasets, respectively, with 5.2 million parameters and minimal computational complexity. Ultimately, the model attained 94.00% accuracy on the novel soybean leaf dataset [10].

Table 1. Compare this work with the related recent work

Author	Dataset Size	Target Plant Disease	Advantage	Limitations
S. Raveena., et al, [11]	640 images	Coffee Berry Disease (Colletotrichum kahawae)	High accuracy (98.56%) was achieved using XG-Boost and 10-fold cross-validation. Identifies key features like cherry color, spots, and lesion size.	Requires a large and diverse dataset to avoid overfitting. Sensitivity to noise in the data.
A. Mohamed., et al, [12]	642 images	Coffee leaf rust, Coffee leaf miner	Utilizes multiple convolutional filters of different sizes in parallel for efficient feature extraction.	Requires more parameters and computational resources compared to some other models.
M. Nawaz., et al, [13]	18,985 images	Coffee plant leaf diseases	High accuracy (98.54% classification accuracy and 0.97 mAP). A novel approach combining CenterNet with spatial-channel attention and ResNet-50	Severe disease symptoms were not encountered in the training data.

E.D. Nugroho., et al, [14]	1860 coffee bean images	Robusta coffee bean defects: Broken, Black, Wrinkled, and Moldy/Bleached	High detection accuracy for some defects (e.g., black beans: 95.3%).	Lower accuracy for certain defects (e.g., moldy, bleached beans: 62.2%)
A. Upadhyay., et al, [15]	54,306 photos of 14 distinct crops representing 26 plant diseases	Diverse plant diseases (e.g., leaf spot, rust, powdery mildew, etc.)	Non-destructive, fast, real-time, and precise disease detection. Minimizes human bias in feature selection.	Required High-quality datasets for optimal model training.
S. Thandapani., et al, [16]	1058 images	Cinnamon plant	Employs two feature extraction levels (LDPP and LTCP) for improved feature representation.	Relies on traditional machine learning models, which are less effective than deep learning approaches for complex image analysis.

Materials and Methods

To implement an object detection and disease prediction system using YOLOv5, including visual processing and deep learning methodologies, for the real-time and precise recognition of objects and forecasting of Coffee Berry Blotch diseases.

Dataset and Preprocessing

The initial vital stage in developing a high-quality dataset for object identification and disease prediction involves collecting images. Accurate image capture improves the dataset's efficacy and variety, hence improving model precision and robustness. By using high-resolution cameras or drones equipped with macro lenses to get precise images of coffee berries. Cameras are strategically positioned at varying distances and angles in the field to capture berries in their natural habitat. Either hand-held or fixed cameras can be used based on the circumstances. Environmental factors concerning lighting settings to guarantee model generality include capturing images in strong sunshine, overcast circumstances, and artificial illumination. Diverse backdrops of photos illustrating branches, leaves, dirt, and other botanical components. Analyze berries at various stages of ripeness. Data Collection Comprises 1875 high-resolution photos (224x224 pixels) of berries impacted by berry blotch disease at various phases (Healthy Berries, Mildly Infected, Moderately Infected, and Severely Infected) Figure 1. Captured photographs of berries in clusters (many berries on a branch) as well as solitary berries. Image normalization is scaling pixel intensity values to a predefined range (0-1 or -1 to 1) to enhance model stability and convergence. Standardization involves subtracting the mean and dividing it by the standard deviation of pixel values across the collection. Data augmentation is performed using Tensor-Flow's Image Data Generator to automate operations such as rotating photos at various angles (90°, 180°, 270°) to simulate varied orientations. Implementing horizontal and vertical flips to enhance data variety. Adjusting scale to simulate varying camera distances. Brightness and contrast are modified accordingly to replicate fluctuations in natural lighting situations. Gaussian noise introduces little disturbances to pictures, enhancing the model's robustness against flaws in real-world applications.

Figure 1. Sample Dataset Collection of Coffee Berry Blotch Disease

Ultimately, verify that the supplemented data preserves an equilibrium between healthy and unhealthy berries. Enhance underrepresented categories (infrequent illness stages) to get a balanced dataset.

Object Detection using YOLOv5

The input images contain high-resolution images of coffee berries. The images consist of healthy berries, diseased berries (affected by blotch), and background features (branches, healthy berries, foliage). The images undergo preprocessing (resizing to a predetermined dimension, normalization, etc.) before being input into the YOLOv5 Framework, as shown in Figure 2. Feature extraction acts as a framework for deriving features from input images Eq 1. It employs convolutional layers to identify patterns like edges, textures, and advanced features.

$$\mathrm{F} = \mathrm{ReLU}\big(\mathrm{BN}(\mathrm{W} * \mathrm{I}' + \mathrm{b})\big) \quad (1)$$

Where W and b are the weights and biases of the convolutional layer, $*$ represent the convolution operation, BN is batch normalization, and ReLU is the activation function.

$$F_{CSP} = Concat(F_1, F_2) \quad (2)$$

Where F_1 and F_2 are features processed in parallel branches of the CSP module. Bottle Neck CSP (Cross-Stage Partial Networks) Eq 2, begins with multiple Bottle Neck CSP modules, which are efficient convolutional structures. They partition the feature map, process it, and then recombine it to enhance gradient flow and minimize computing load. These modules extract multi-scale characteristics from coffee berry images, including spots, blotches, and fruit outlines. SPP (Spatial Pyramid Pooling) module captures spatial information by pooling the feature map at different scales (max pooling with varying kernel sizes) Eq 3 Where k is the pooling kernel size which enhances the model's ability to detect blotch patterns regardless of the size or position of the diseased area on the coffee berries.

F_SPP =Concat(MaxPool_(k=1) (F),MaxPool_(k=5) (F),MaxPool_(k=9) (F),MaxPool_(k=13) (F)) (3)

PANet (Path Aggregation Network) integrates features from multiple levels to enhance detection accuracy. Features from the backbone are upsampled (Features from higher layers are upsampled using interpolation) Eq 4, and concatenated (Upsampled features are concatenated with features from earlier layers) Eq 5, with feature maps.

$$F_{up} = Upsample(F_{SPP}) \quad (4)$$
$$F_{PANet} = Concet(F_{Up}, F_{Low}) \quad (5)$$

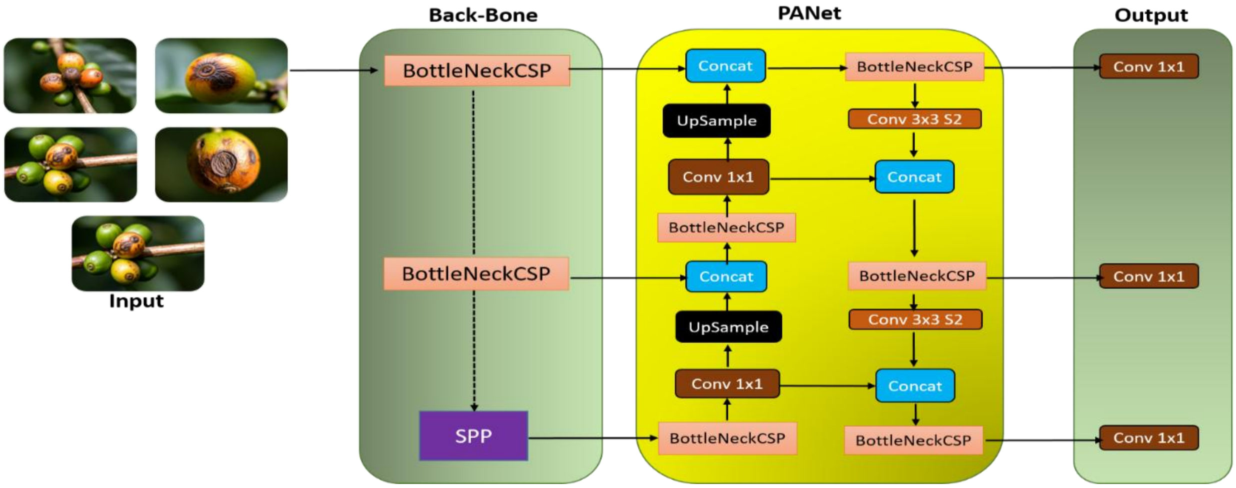

Figure 2. Structure of YOLOv5 Framework

This integration enables the model to combine low-level (fine-grained) and high-level (semantic) information. For example, fine-grained characteristics will help in identifying the boundaries of a blotch, while high-level features concentrate on the overarching diseased pattern. Convolutional layers Eq 6, use several convolutional operations (1x1 and 3x3) to enhance the aggregated features. The objective is to improve the network's comprehension of the spatial configuration and patterns present in the images.

$$F_{refined} = ReLU\big(BN(W * F_{PANet} + b)\big) \quad (6)$$

The output layer's enhanced feature maps from the PANet are sent to the detection head, where bounding boxes are predicted for the regions of interest (ROIs). Each bounding box has coordinates (x, y, width, height) Eq 7, where, t_x, t_y, t_w, t_h are predicted offsets, c_x, c_y c_x, c_y are the cell offsets, p_w, p_h, p_w, p_h are the anchor box dimensions, and σ is the sigmoid activation function. Confidence score (the likelihood that the box contains an item) Eq 8 where C is the confidence logit. Class name (e.g., "Healthy", "Mildly Diseased", "Moderately Diseased", "Severely Diseased") Eq 9, where z is the output of the final convolutional layer. YOLOv5 employs multi-scale detection to recognize items at three distinct scales, enhancing its efficacy in recognizing both small and large diseased regions.

$$x = \sigma(t_x) + c_x, \qquad y = \sigma(t_y) + c_y$$
$$w = p_w e^{tw}, h = p_h e^{th} \quad (7)$$
$$P_{conf} = \sigma(C) \quad (8)$$
$$P(class) = Softmax(z) \quad (9)$$

Loss functions are evaluated during training The total loss (L) in YOLOv5 is composed of three main components of the loss (L_{box}, L_{cls}, L_{obj}) and can have different scales. λ_{box} ensures the bounding box loss contributes appropriately to the total loss. Eq 10 where λ_{box} is a weighting factor.

$$L = \lambda_{box} L_{box} + \lambda_{obj} L_{obj} + \lambda_{cls} L_{cls} \quad (10)$$

Bounding Box Loss (L_{box}) assesses the accuracy of the predicted bounding box about the ground truth. YOLOv5 employs Complete Intersection over Union (CIoU) loss to enhance bounding box regression.

$$L_{box} = 1 - CIoU\big(B_p, B_{gt}\big) \quad (11)$$

Where Eq 11, B_p: Predicted bounding box, B_{gt}: Ground truth bounding box, and CIoU incorporates overlap, center distance, and aspect ratio. Objectness Loss (L_{obj}) assesses the presence of an object inside a specified grid cell. Binary cross-entropy (BCE) loss is used.

$$L_{obj} = -\sum_{j=1}^{m}\left[y_j \log(p_j) + (1 - y_j)\log(1 - p_j)\right] \tag{12}$$

Where Eq 12, y_i: Ground truth Objectness (1 for presence, 0 for absence), and p_i: Predicted Objectness probability. Classification Loss (L_{cls}) measures how well the predicted class matches the ground truth and uses binary cross-entropy for multi-label classification

$$L_{cls} = -\sum_{j=1}^{m}\left[y_j \log(p_j) + (1 - y_j)\log(1 - p_j)\right] \tag{13}$$

Where Eq 13 illustrates M: Number of classes, y_j: Ground truth class label (1 if true, 0 otherwise), and p_j: Predicted class probability.

Disease Classification and Prediction using Hybrid Vision Transformer (ViT)-CNN Models

An advanced hybrid architecture integrating Vision Transformer (ViT) and Convolutional Neural Network (CNN) for the effective classification and prediction of diseases in coffee berries. High-resolution images of coffee berries (both healthy and diseased) are fed into the model. Vision Transformer (ViT) Framework patch partitioning involves dividing the image provided into fixed-size segments, such as 16×16 pixels. Each patch is converted into a vector for further processing. In linear embedding, flattened patches undergo a linear projection layer to be embedded into a higher-dimensional space, resulting in patch embedding. Positional encoding is used to maintain spatial links across patches by including positional encodings into the embedding. Transformer encoding employs multi-head self-attention and feed-forward layers, enabling the Vision Transformer to store global dependencies and contextual interactions across the whole image. The output of this path is a feature map (F_{ViT}) representing the global context of the image. Path of Convolutional Neural Network (CNN) Convolutional layers apply several convolutions on the same input images. These layers implement filters (e.g., 32×3×3) to extract local characteristics such as textures, edges, and patterns. Downsampling using pooling layers reduces spatial dimensions, preserving critical characteristics while lowering computing complexity. In up-sampling after down-sampling, feature maps are enlarged to ensure resolution compatibility with the outputs of the ViT. The CNN path generates a feature map (F_{CNN}) that captures local patterns critical for disease detection. Attention-based Fusion Mechanism as illustrated in Figure 3, global features (F ViT) and local features (FCNN) are integrated via an attention method Eq 14. The attention mechanism allocates weights to highlight essential characteristics from both pathways, guaranteeing that the fused features (F fused) capitalize on the advantages of both ViT (global) and CNN (local) representations.

$$F_{fused} = Attention\ (F_{ViT}, F_{CNN}) \tag{14}$$

The fused feature map is passed through fully connected (dense) layers of the classification head and the final layer uses a softmax activation function to output probabilities for each disease class where, outputs include: Healthy Berry, Mildly Diseased, Moderate Diseased, Severely Diseased).

Experimental Results and Discussions

Employing Yolov5 and Vision Transformer-based Convolutional Neural Network analysis to evaluate the proposed model's performance in terms of accuracy and loss.

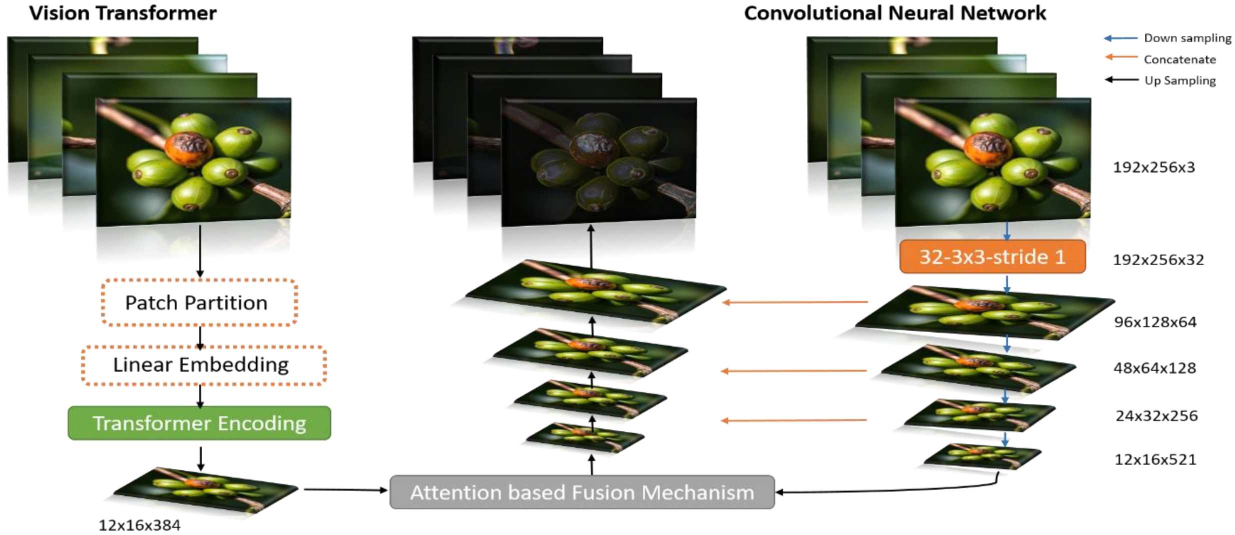

Figure 3. Structure of Vision Transformer-based CNN Model

Evaluation Setup

Python 3.6.5 implements the proposed model on a PC equipped with an i5-8600k processor, GeForce 1050Ti 4 GB graphics card, 16 GB of RAM, a 250 GB SSD, and a 1 TB HDD. The suggested technique seems to surpass others for an image size of 224 × 224 × 3. The proposed model is assessed using True Positive, True Negative, False Negative, and False Positive metrics by Yolov5 and Vision Transformer-based CNN analysis. The parameter configurations are as follows: Epochs: 45, learning rate: 0.01, batch size: 5, dropout: 0.5. The proposed model was developed using the Keras Python framework and library version 2.7. The experiments included input picture dimensions ranging from 192x256 by 3 to 12x16 by 521.

Performance Evaluation for Object Detection Metrics

Detecting objects in disease prediction Figure 4, such as identifying diseased berries, emphasizes the assessment of bounding box localization and classification as illustrated in Table 2.

Table 2. Performance Metric for Object Detection

Sample	Entire Images		Precision	Recall	IoU	mAP		Precision	Recall	IoU	mAP
Healthy Berry	269		0.92	0.92	0.90	0.89		0.96	0.95	0.94	0.92
Mildly Diseased	467	YOLO v3	0.90	0.88	0.85	0.87	YOLO v5	0.92	0.94	0.91	0.93
Moderate Infection	562		0.92	0.89	0.87	0.88		0.96	0.93	0.91	0.94
Severely Infected	577		0.91	0.90	0.86	0.89		0.93	0.95	0.91	0.94
Total Samples	1875		0.92	0.89	0.87	0.88		0.94	0.93	0.96	0.94

$$P = (True\ Positive\ (TP))/(True\ Positive(TP) + \big(False\ Positive(FP)\big) \tag{15}$$

$$R = (True\ Positive\ (TP))/(True\ Positive(TP) + (False\ Negative(FN))) \quad (16)$$

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \quad (17)$$

$$mAP\ = \left(\frac{1}{N}\right) * \ \Sigma(APi) \quad (18)$$

Prevalent metrics encompass: Precision Eq 15, Recall Eq 16, Intersection over Union Eq 17, and Mean Average Precision (mAP) Eq 18.

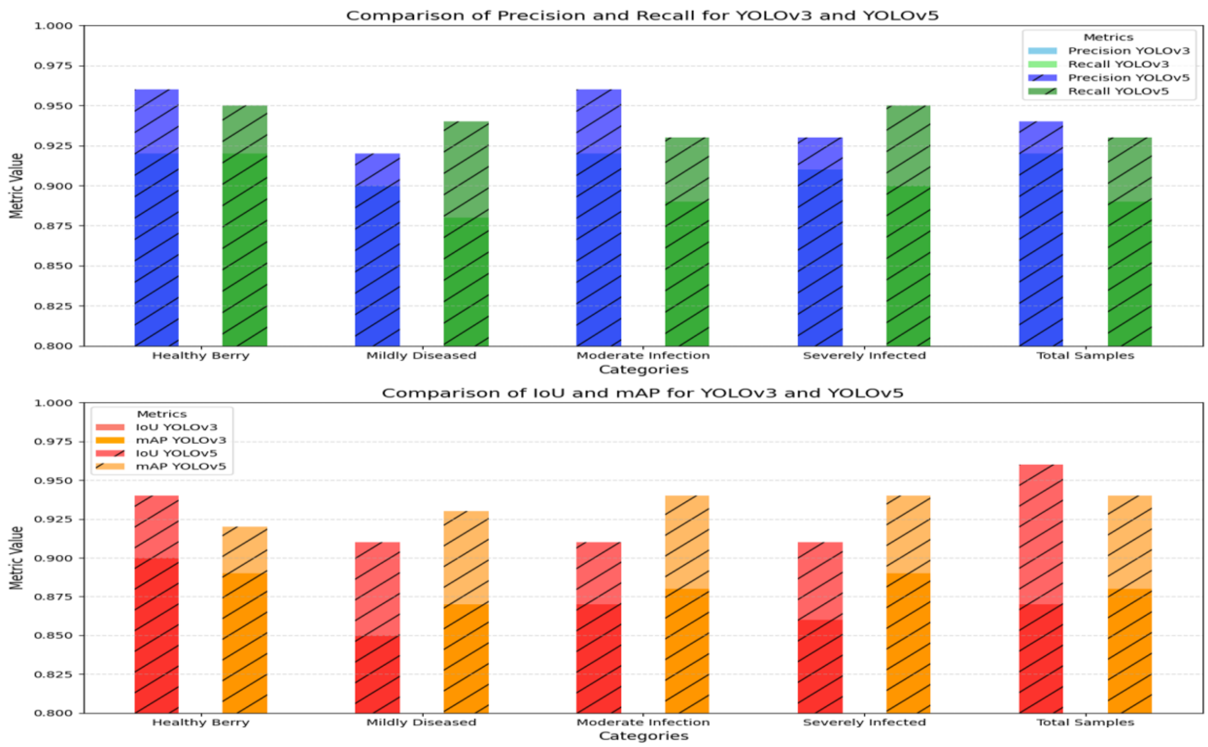

Figure 4. Performance Evaluation over for YOLOv3 and YOLOv5

Performance Evaluation for Disease Detection and Classification Metrics

The classification and prediction as illustrated in Figure 5, of diseases depend on the model's capacity to properly detect and classify them as illustrated in Table 3. Metrics include Accuracy Eq 19, F1-Score Eq 20.

Table 3. Performance Metric for Disease Classification and Prediction

Sample	Entire Images	Convolutional Neural Networks				Vision Transformer based CNN			
		Precision	Accy	Recall	F1-Score	Precision	Accy	Recall	F1-Score
Healthy Berry	269	85.0%	89.0%	80.0%	82.4%	92.0%	94.5%	90.0%	91.0%
Mildly Diseased	467	88.2%	96.2%	86.0%	87.1%	93.5%	96.2%	91.7%	92.6%
Moderate Infection	562	86.5%	90.7%	84.5%	85.5%	91.8%	95.0%	89.9%	90.8%
Severely Infected	577	84.8%	88.0%	82.0%	83.4%	90.7%	94.0%	88.5%	89.6%
Total Samples	1875	86.1%	**89.8%**	83.1%	84.6%	92.0%	**95.7%**	90.0%	91.0%

$$Accuracy = \frac{Number\ of\ Correct\ Prediction}{Total\ Number\ of\ Prediction} \tag{19}$$

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{20}$$

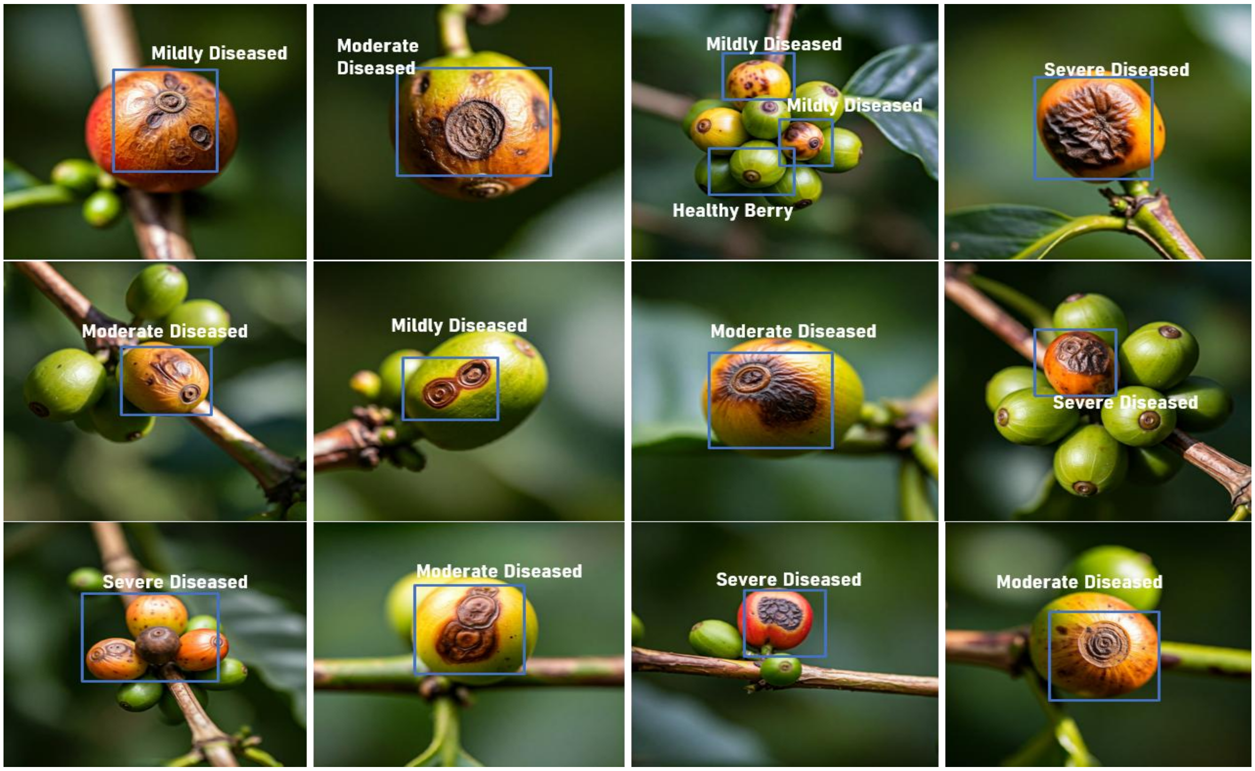

Figure 5. Collected Samples Are Classification and Prediction

Conclusions

This research examines the crucial challenge of identifying and evaluating the severity of berry blotch disease in coffee plantations, which directly affects crop quality and yield. The objective is to provide a comprehensive, automated solution to aid coffee producers in efficiently monitoring plant health. A vision-based object identification method using YOLOv5 and hybrid ViT-CNN models was executed to identify and estimate the severity of berry blotch disease. The proposed approach utilizes deep learning to analyze high-resolution images from coffee plantations. The algorithm achieved an accuracy of 95.7%, demonstrating its efficacy in detecting berry blotch disease and accurately assessing severity levels. The method has the potential to improve early disease identification and facilitate informed decision-making. The model's performance can also be influenced by ambient factors, like lighting fluctuations and obstructions. Future efforts should concentrate on augmenting the dataset, enhancing model resilience, and incorporating IoT-based real-time monitoring for practical implementation.

References

1. O. Adelaja, and B. Pranggono, “Leveraging Deep Learning for Real-Time Coffee Leaf Disease Identification”, *Agri Engineering*, vol. 7, no. 1, pp. 13, 2025. https://doi.org/10.3390/agriengineering7010013

2. X. Wang, C. Zhang, Z. Qiang, C. Liu, X. Wei, and F. Cheng, “A Coffee Plant Counting Method Based on Dual-Channel NMS and YOLOv9 Leveraging UAV Multispectral Imaging”, *Remote Sensing*, vol. 16*,* no. 20, pp.3810, 2024.
https://doi.org/10.3390/rs16203810
3. P.S. Reddy, R. Surendran, K. Divya, K., R. Selvanarayanan, S.K. Selvaperumal, and R. Lakshamanan, “Accuracy analysis and comparison of crop yield prediction using NB algorithm and RNN”, *In AIP Conference Proceedings AIP Publishing,* Vol. 3161, No. 1, pp. 020363, 2024.
https://doi.org/10.1063/5.0229471
4. R. Li, Y. He, Y. Li, W. Qin, A. Abbas, R. Ji, S. Li, Y. Wu, X. Sun, and J. Yang, “Identification of cotton pest and disease based on CFNet-VoV-GCSP-LSKNet-YOLOv8s: a new era of precision agriculture”, *Frontiers in Plant Science*, vol. 15, pp. 1348402, 2024.
https://doi.org/10.3389/fpls.2024.1348402
5. R. S and S. R, "Clustering-based Hemileia Vastatrix Disease Prediction in Coffee Leaf using Deep Belief Network", *8th International Conference on Communication and Electronics Systems (ICCES),* vol. 2, pp. 1094-1100, 2023.
doi: 10.1109/ICCES57224.2023.10192835.
6. C. B. Nwaneto, C. Yinka-Banjo, and O. Ugot, “An object detection solution for early detection of taro leaf blight disease in the West African sub-region”, *Franklin Open*, vol. 10, pp. 100197, 2025.
https://doi.org/10.1016/j.fraope.2024.100197
7. A. Pavate, S. Kukreja, S. Janrao, S. Bankar, R. Patil, and V. Bidve, “Efficient model for cotton plant health monitoring via YOLO-based disease prediction”, *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 1, pp. 164-178, 2025.
https://doi.org/10.11591/ijeecs.v37.i3
8. G. Polder, P.M. Blok, T. van Daalen, J. Peller, and N. Mylonas, “A Smart Camera With Integrated Deep Learning Processing for Disease Detection in Open Field Crops of Grape, Apple, and Carrot”, *Journal of Field Robotics*, vol. 1, pp. 1-14, 2025.
https://doi.org/10.1001/rob.22510
9. C. A. Elinisa, C. Wa Maina, A. Vodacek, and N. Mduma, “Image Segmentation Deep Learning Model for Early Detection of Banana Diseases”, *Applied Artificial Intelligence*, vol. 39, no. 1, pp. 2440837, 2025.
https://doi.org/10.1080/08839514.2024.2440837
10. V. Sharma, A.K. Tripathi, H. Mittal, and L. Nkenyereye, “SoyaTrans: A novel transformer model for fine-grained visual classification of soybean leaf disease diagnosis”, *Expert Systems with Applications*, vol. 260, pp.125385, 2025.
https://doi.org/10.1016/j.eswa.2024.125385
11. S. Raveena, S. Rajendran, and Y. Alotaibi. "Early Detection of Colletotrichum Kahawae Disease in Coffee Cherry Based on Computer Vision Techniques", *CMES-Computer Modeling in Engineering & Sciences,* vol. 139, no. 1, pp. 759-782, 2024.
12. A. Mohamed, and N. N. Mostafa, "Deep Learning for Coffee Leaf Diseases Detection in Precision Agriculture", *Optimization in Agriculture,* vol. 1, pp. 129-136, 2024.
https://doi.org/10.61356/j.oia.2024.1292

13. M. Nawaz, T. Nazir, A. Javed, S. T. Amin, F. Jeribi, and A. Tahir, "CoffeeNet: A deep learning approach for coffee plant leaves diseases recognition", *Expert Systems with Applications*, vol. 237, pp. 121481, 2024.
https://doi.org/10.1016/j.eswa.2023.121481
14. E.D. Nugroho, M. Verdiana, M.H. Algifari, A. Afriansyah, H.B. Firmansyah, A.K. Rizkita, R.A. Winarta, and D. Gunawan, "Development of YOLO-Based Mobile Application for Detection of Defect Types in Robusta Coffee Beans", *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 153-160, 2025.
https://doi.org/10.30871/jaic.v9i1.8886
15. A. Upadhyay, N.S. Chandel, K. P. Singh, S.K. Chakraborty, B.M. Nandede, M. Kumar, A. Subeesh, K. Upendar, A. Salem, and A. Elbeltagi, "Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture", *Artificial Intelligence Review*, vol. 58, no. 3, pp. 1-64, 2025.
https://doi.org/10.1007/s10462-024-11100-x
16. S. Thandapani, D. Selvaraj, M.I. Mahaboob, V. C. Bharathi, and P. Vinayagam, "Swift and efficient cinnamon plant disease classification using robust feature extraction and machine learning techniques", *Bulletin of Electrical Engineering and Informatics*, vol. 14, no. 1, pp. 421-428, 2025.
https://doi.org/10.11591/eei.v14i1.8254
17. L. Ghiani, S. Serra, A. Sassu, A. Deidda, A. Deidda, and F. Gambella, "Automated Detection of Downy Mildew and Powdery Mildew Symptoms for Vineyard Disease Management", *Bulletin of Electrical Engineering and Information*, vol. 14, no. 1, pp. 421-428, 2025.
https://doi.org/10.11591/eei.v14i1.8254
18. R. Selvanarayanan, and S. Rajendran, "Roaming the Coffee Plantations Using Grey Wolves Optimisation and the Restricted Boltzmann Machine to Predict Coffee Berry Disease", *International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, pp. 681-689, 2023.
https://doi.org/10.1109/ICSSAS57918.2023.10331629
19. L. Ghiani, S. Serra, A. Sassu, A. Deidda, A. Deidda, and F. Gambella, "Automated Detection of Downy Mildew and Powdery Mildew Symptoms for Vineyard Disease Management", vol. 2, pp. 1-33, 2025.
http://dx.doi.org/10.2139/ssrn.5085231
20. W. Yuan, L. Lan, J. Xu, T. Sun, X. Wang, Q. Wang, J. Hu, and B. Wang, "Smart Agricultural Pest Detection Using I-YOLOv10-SC: An Improved Object Detection Framework", *Agronomy*, vol. 15, no. 1, pp. 221, 2025.
https://doi.org/10.3390/agronomy15010221
21. S. Raveena, and R. Surendran, "ResNet50-based classification of coffee cherry maturity using deep-CNN", *In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)* vol. 1, pp. 1275-1281, 2023.
https://doi.org/10.1109/ICSSIT55814.2023.10061006