International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

A Smart Agriculture Framework for Apple Leaf Disease Detection Using YOLOv8 and Swin Transformer Networks

[1]Pritam Kumar Gupta, [2]Prof. Kamlesh Raghuwanshi, [3]Prakash maravi, [4]Dr.Surabhi Karsoliya
[1,2,3,4]Department of Artificial Intelligence & Machine Learning, Technocrats Institute Of Technology(excellence), Bhopal (M.P.)
[1]54321gupta.pritam@gmail.com , [2]raghu.kamlesh82@gmai.com, [3]prkshmaravi1@gmail.com , [4]surabhi.karsoliya@gmail.com

ABSTRACT

Detecting diseases in apple leaves plays a vital role in maintaining agricultural productivity, output quantity, and product standards. Enhancing the quality of cultivated crops and increasing their yield can be achieved by reducing losses and adopting better disease management strategies through effective disease diagnosis. Currently, the primary approach to diagnosing plant diseases involves conventional techniques, including the examination of leaves and branches by an experienced professional. This process is not only lengthy and expensive but also prone to inaccuracies due to human disruptions. Over the years, learning methods have developed and shown notable progress in enhancing techniques to combat plant diseases. This research differs primarily in utilizing the YOLOv8s method and a hybrid ResNet50 plus Swin Transformer model with attention features to target the classification of apple leaf diseases. The method enhances performance in object decomposition and disease classification by leveraging Transformer models for global context and CNNs for extracting local features. With a total accuracy of 98.56%, the hybrid system also achieved a weighted average F1-score of 98.56% and a macro average F1-score of 98.59%. Since disease category A achieved an F1-score of 99.66%, it suggests that hybrid plant disease monitoring is highly effective in managing Cedar Apple Rust specifically. The confusion matrix and the precision-recall analysis offered additional evidence supporting this highly uniform classification, which exhibited almost no errors across the different disease categories. The proposed hybrid framework achieved improved classification accuracy, enhanced feature extraction, and superior prediction performance relative to the current YOLOv8s model. The suggested hybrid framework obtained strong classification performance by capturing both basic spatial features and more complex spatial patterns with contextual details, despite YOLOv8s being trained as an end-to-end model that performed well in localization and detection tasks.

Keywords: Apple Leaf Disease Detection, YOLOv8s, Hybrid ResNet50 + Swin Transformer, Deep Learning, Smart Agriculture , CNN, Hybrid, disease

I. INTRODUCTION

Apple farming has a major part in worldwide agriculture because apples contain high nutritional value , they matter commercially, and they bring solid economic returns to the horticulture market. But apple production is still heavily influenced by several leaf diseases like Apple Scab, Cedar Apple Rust, Powdery Mildew, and Black Rot, and these problems lower the crop quality , reduce the harvest size, and hurt farmers financial outcomes. Because of this, detecting sickness early and correctly becomes very important for better crop productivity and for keeping agricultural development more sustainable. Traditionally, people diagnose these leaf diseases by direct manual inspection from agricultural specialists. This way is often slow, labor-demanding, costly, and when orchards are large it can become unreliable. In addition , everyday environmental conditions make the whole task harder, for example changing illumination, leaf occlusion, complex backgrounds, and overlapping foliage [1]. These factors limit the efficiency of usual disease surveillance approaches in real orchards. Recent advancements in Artificial Intelligence , Deep Learning and Computer Vision have been changing agricultural monitoring, by enabling smarter and automated disease detection frameworks. In many cases, Convolutional Neural Networks and object detection models show strong results when recognizing plant illnesses from leaf images. Within these methods, the YOLO architecture, You Only Look Once, has pulled a lot of attention, mainly because it can do real-time object detection , runs efficiently and tends to localize objects very well. In particular YOLOv8 gives better feature extraction, anchor-free detection, and improved recognition of tiny targets , so it is a good fit for apple leaf disease identification even when orchards are complex and visually messy. In addition to YOLO based models [2], Transformer setups like Swin Transformer have been getting real traction for image analysis lately, mainly because they grab long range dependencies and also pull in overall contextual cues from an image. Compared to standard CNNs, which tend to emphasize local patterns, Swin Transformer helps disease recognition by looking at how regions relate to each other across the full scene. So, if YOLOv8 is combined with Swin Transformer networks, the result should improve disease localization, boost the quality of the learned feature maps, and also raise classification accuracy at the same time [3]-[4]. This research paper propose a smart agriculture framework for apple leaf disease recognition using YOLOv8 ,and Swin Transformer based networks. The framework is intended to deliver precise , real time and automated disease identification for precision farming tasks. By weaving in advanced deep learning methods it can support adaptive agricultural monitoring, reduce crop losses, enhance disease management efficiency and help move toward more sustainable smart farming systems.

II. RELATED WORK

Raj et al. [1] presented YOLO-ODD, which is actually an upgraded YOLOv8s style framework, aimed at detecting onion foliar disease. In their design, attention mechanisms were added in a way that helped the model not only see more clearly but also localize better, so identification stayed reliable even when the field conditions were messier than expected. In experiments, the method showed higher precision, recall, and overall robustness versus the more conventional YOLOv8s setups, which is why the authors highlight it as a meaningful step. Then Bao and Zhang [2] came up with an apple pest and disease detection network built around partial multi scale feature extraction plus an efficient hierarchical feature fusion scheme. The overall intent was to strengthen the recognition of small disease patches, as well as those tricky pests, especially when images were taken in orchard settings where background clutter happens often. Their findings suggested better feature representation, and they also reported improved disease detection accuracy compared with baseline approaches. Zarboubi et al. [3] came up with an IoT-integrated, YOLO-Evo framework for more sustainable apple farming use. In their idea the system mixes IoT sensors and a YOLO- based object detection approach, for, intelligent pest monitoring and real time watching of orchards. In the end it gave cost effective and automated agricultural monitoring results. Li et al. [4] suggested HEFM-YOLO, which is a lightweight apple leaf disease detection model made for everyday farm conditions. Their design drops Hybrid Enhanced Feature Modules into the YOLO structure to help with better disease localization and richer feature extraction. The experiments showed strong accuracy while keeping efficient real time behavior. Erkamim et al. [5] looked at various YOLOv5 convolutional neural network setups for detecting apple leaf disease. They tested how well different YOLOv5 versions work in real orchard scenarios and the results show faster recognition of the symptoms , and also higher precision, which supports precision agriculture workflows. Zhang et al. [6] presented YOLO-ACT, this adaptive cross-layer integration approach aimed at apple leaf disease detection. Their idea strengthens the interaction among features across several layers , so the system becomes better at both finding where the disease is and staying reliable even when illumination changes a bit. Li et al. [7] created YOLO-Leaf, a more capable deep learning framework for apple leaf disease identification in precision agriculture platforms. The design improves the object detection pipeline along with the feature extraction process , and this leads to strong real time disease detection, plus better classification outcomes overall. Yan and Yang [8] proposed FSM-YOLO, it is an apple leaf disease detection network that uses adaptive feature capture and a sense of spatial context, more or less. In the framework they brought in attention mechanisms and modules for adaptive feature extraction, this helps disease recognition a lot, especially in hard agricultural settings Yuan et al. [9] put forward a revised method built on YOLOv7, aimed at apple leaf disease detection. They tuned feature extraction and disease localization parts within the YOLOv7 pipeline, which resulted in strong precision and recall too, while keeping disease recognition efficient, without lag Gomez et al. [10] looked into YOLO-driven deep learning models, targeted for common bean disease detection, in farm environments. Their findings showed that YOLO based approaches meaningfully raised detection accuracy and also strengthened real time agricultural monitoring, overall performance stayed practical. Yang et al. [11] came up with CA-YOLOv5 which is a coordinate attention based YOLO setup, it targets apple detection in natural orchard settings. In their design, localization accuracy got better when fruits were blocked or lighting changes happened, and that ended up boosting overall detection performance for smart agriculture systems. Reim et al. [12] built a phenotyping workflow grounded on YOLO, meant for apple blotch disease evaluation after artificial inoculation. Their approach helped with accurate symptom recognition and also made automated agricultural phenotyping tasks easier to run without too much manual effort. Boudaa et al. [13] looked into pre-trained YOLO models for plant disease detection jobs. They showed that transfer learning with pretrained architectures can really raise training efficiency, while also improving disease recognition accuracy for agricultural image analysis. Sangaiah et al. [14] came up with UAV T-YOLO-rice, which is basically a improved Tiny-YOLO scheme for spotting rice leaf illness from UAV pictures. In their setup, disease recognition ran fast inside paddy plots and, overall , it showed the value of UAV driven intelligent farming systems. Zhou et al. [15] presented YOLO-AppleSeg, a slim instance segmentation network meant for apple fruit location and tallying. Their method produced solid masks and also kept the detection procedure efficient when used in orchard scenes. Xiao et al. [16] used a Transformer based visual analysis pipeline for judging apple ripeness through digital images. They pointed out that Transformer designs can gather wider contextual cues and this lifted the classification accuracy for agricultural imaging tasks. Sapkota et al. [17] looked into how LLM generated synthetic datasets can be used with YOLO11 and YOLOv10 for apple detection systems. The work indicated that when you do synthetic data augmentation , it can boost how well the model generalizes and it boosts detection capability in machine vision use cases. Anitha et al. [18] came up with a YOLOv3 convolutional neural network setup for pest detection plus identification , focused on rice crops. The model showed efficient real time pest monitoring , so it fits agricultural surveillance setups that need continuous observation. Hu et al. [19] presented BHI-YOLO, which is a lightweight instance segmentation approach for strawberry disease detection. Their framework supported precise disease localization with low computational load. Because of that, it fits edge computing scenarios quite well, even with limited resources. BalaChandralekha and Thangakumar [20] worked on a YOLOv8 deep learning framework for detecting fungal disease in apple plants, it was focused on better disease identification. Their proposed model showed strong precision and recall, plus efficient recognition of the disease under smart agriculture settings, and it was used for field like conditions too.

III. RESEARCH METHODOLOGY

A. Dataset Collection

The apple leaf dataset was collected from publicly available agricultural image repositories. The dataset contained four categories of apple leaves, namely Cedar Apple Rust, Apple Scab, Black Rot, and Healthy leaves. The images were captured under different environmental conditions including varying illumination, background complexity, and leaf orientations to improve model robustness and real-world applicability.

B. Image Preprocessing

Image preprocessing techniques were applied to improve image quality and prepare the dataset for deep learning analysis. The preprocessing stage included image resizing, normalization, noise removal, and contrast enhancement. Data augmentation techniques such as rotation, flipping, scaling, cropping, and affine transformation were also applied to increase dataset diversity and reduce overfitting during model training.

C. YOLOv8-Based Disease Detection

The YOLOv8 model was utilized for automated apple leaf disease detection and localization. YOLOv8 is a single-stage object detection framework capable of predicting bounding boxes and disease categories simultaneously. The model was selected due to its fast inference speed, anchor-free detection mechanism, improved feature extraction capability, and superior localization performance. The framework effectively identified infected regions on apple leaves under complex orchard conditions.

D. Hybrid ResNet50 and Swin Transformer Classification

After disease localization, the detected leaf regions were passed to a Hybrid ResNet50 and Swin Transformer classification framework. ResNet50 extracted local spatial and texture features from diseased leaf images using convolutional operations and residual learning. Simultaneously, the Swin Transformer captured global contextual information and long-range dependencies through self-attention mechanisms. The extracted features were fused to generate robust disease representations for accurate classification. Figure 1 illustrates the proposed Hybrid Attention-Based ResNet50–Swin Transformer framework integrated with Bayesian Optimization for accurate and efficient apple leaf disease classification in smart agriculture applications.

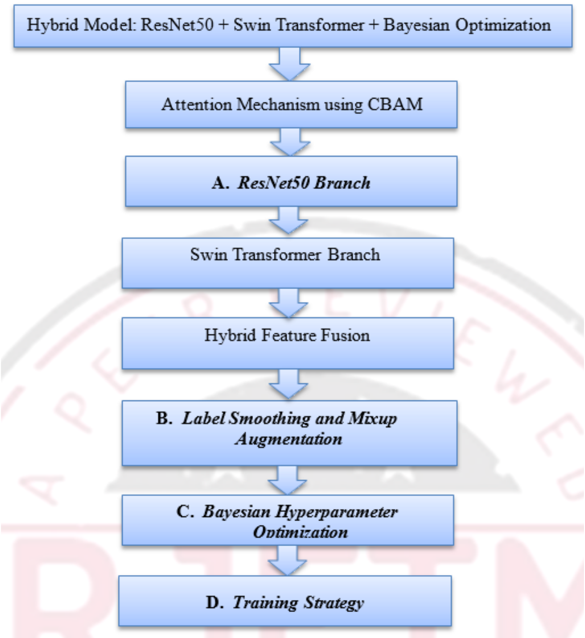


Figure 1: Hybrid Attention-Based ResNet50–Swin Transformer Architecture with Bayesian Optimization for Apple Leaf Disease Classification

E. Model Training and Optimization

The proposed framework was trained using transfer learning techniques and optimized using the Adam optimizer with categorical cross-entropy loss. Hyperparameters including learning rate, batch size, and confidence threshold were adjusted to improve training stability and detection accuracy. Batch normalization and dropout layers were also incorporated to reduce overfitting and improve model generalization capability.

F. Performance Evaluation

The performance of the proposed framework was evaluated using precision, recall, F1-score, accuracy, Intersection over Union (IoU), mean Average Precision (mAP@0.5), and mAP@0.5:0.95 metrics. Confusion matrix analysis and ROC curve evaluation were also performed to analyze classification consistency and disease recognition capability.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Precision is a measure of the ability of a model to make correct positive disease predictions. It assesses the proportion of leaf images classified as diseased that are from the diseased leaf category.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2}$$

The recall metric in machine learning measures the model’s effectiveness in recognizing true sick leave records from the dataset, emphasizing its capacity to capture real disease cases without missing important instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

The F1-score measures the harmony between Precision and Recall and is particularly valuable when minimizing false positives is a priority.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{4}$$

.Mean Average Precision is the most crucial performance criteria for object detection frameworks such as YOLOv8s. It measures relational accuracy between disease detection and classification capabilities of an image at the same time. Average Precision is calculated by the total area under the precision-recall curve.

$$\text{AP} = \int_1^0 \text{Precision(Recall)dRecall} \tag{5}$$

The mean average precision for the overall study is calculated by averaging the AP values for each illness group.

$$\text{mAP} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i \tag{6}$$

where, N = Total number of disease classes , AP_i = Average Precision of the ith class ,

The calculation of intersection over union is as follows:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{7}$$

The corresponding mAP@0.5 is calculated as:

$$\text{mAP@0.5} = \frac{1}{N}\sum_{i=1}^{N} \text{AP}_i^{\text{IoU}=0.5} \tag{8}$$

This metric measures the AI Tool's capability to effectively identify the extent of the focus in the apple plant. For example, this involves identifying the lesion, rust spots, and discoloration regions on apple leaves. mAP@0.5:0.95 is considered as an advanced detection metric that considers performance in detecting objects over different IoU values including 0.5 to 0.95.

$$\text{mAP@0.5: 0.95} = \frac{1}{10}\sum_{t=0.5}^{0.95} \text{mAP}_t \tag{9}$$

The YOLOv8s framework uses a Fitness score. The features of the score combine the Precision, Recall, and mAP values at the 0.5 threshold.

$$\text{Fitness} = 0.4 \times \text{Precision} + 0.4 \times \text{Recall} + 0.2 \times \text{mAP@0.5} \tag{10}$$

IV. RESULT AND DISCUSSION

This section evaluates the performance of YOLOv8s and Hybrid ResNet50–Swin Transformer models for apple leaf disease detection and classification, demonstrating that the hybrid framework achieved superior accuracy, balanced prediction capability, and effective disease localization for smart agriculture applications.

A. Results and Performance Analysis of Hybrid ResNet50 + Swin Transformer Model

After training, ResNet50 and Swin demonstrated remarkable results, achieving an accuracy of 98.56% on the apple leaf dataset with just one advanced model. The ultimate model demonstrates the most accurate disease detection, as shown in the detailed classification report with exceptionally high precision, recall, and F1 scores across all diseases. **Table 1 presents the final classification report of the proposed Hybrid ResNet50–Swin Transformer model, showing precision, recall, F1-score, and accuracy values for each apple leaf disease category.**

Table 1: Final Classification Report

Class	Precision	Recall	F1-Score	Support
Apple Scab	0.9670	0.9881	0.9774	504
Black Rot	0.9959	0.9738	0.9847	497
Cedar Apple Rust	0.9932	1.0000	0.9966	440
Healthy	0.9880	0.9821	0.9850	502
Overall Accuracy	—	—	0.9856	1943
Macro Average	0.9860	0.9860	0.9859	1943
Weighted Average	0.9857	0.9856	0.9856	1943

Based on the results, the two-phase design showed the least prediction errors and comparable accuracy in classifying certain multimodal clinical groups. Using CNNs for local feature extraction together with transformers for global contextual understanding proves highly effective in classifying this illness, particularly when certain diseases appear very similar. For the Cedar Apple Rust category, the top overall accuracy was indicated by an F1-score of 0.9966 and a recall rate of 1.0000. This confirms that every Cedar Apple Rust sample was correctly identified by the architecture, with no false negatives detected. The effectiveness of the hybrid framework in representing the feature space globally is characterized by high precision and recall. Similarly, Black Rot exhibits an impressive precision of 0.9959 and an F1 score of 0.9847. The accuracy of this prediction is highly refined and precise, resulting in very few false positives or incorrect classifications. Overall, the regular leaf category exhibits a high level of classification accuracy, with precision, recall, and F1-score metrics well above 98%, confirming that the model successfully detects healthy leaves and does not misclassify infected leaves. Compared to other categories, Apple Scab has reduced accuracy, likely because the scab lesions closely resemble rusts or spots in appearance. Nevertheless, with a recall of 0.9881 and an F1-score of 0.9774, it can be concluded that the method's diagnostic effectiveness is highly reliable. The overall dataset achieved an accuracy of 98.6%, indicating that the classes are well-balanced and the detection model performs consistently across all disease types. The summary metrics for recall and the F1-score assist in understanding how well the model can handle changes in class distributions while maintaining strong classification results. The last stage of the classification process shows that the combined ResNet50 and Swin Transformer model effectively merges the traditional CNN's power with transformer-based attention mechanisms, along with Bayesian hyper parameter optimization, resulting in superior and consistent performance in identifying tobacco leaf diseases.

B. Training History of Hybrid ResNet50 + Swin Transformer Model

The Figure 2 above presents the training and validation results for the Hybrid ResNet50 automatically boosted by Swin Transformer during 30 epochs. The convergence, learning stability as well as two classifications are also features of these graphs in the picture. The Loss Curve and Accuracy Curve.

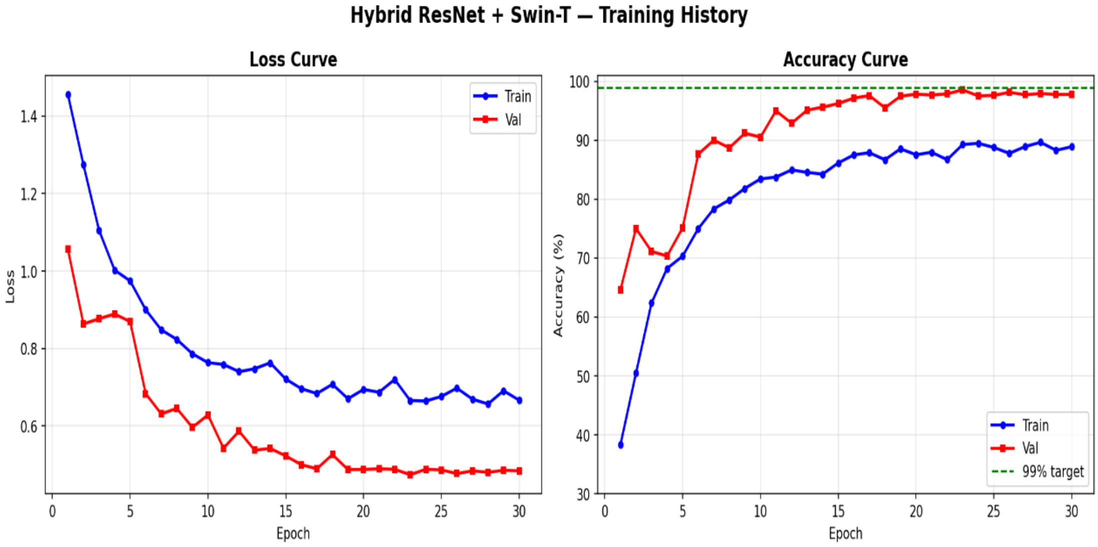


Figure 2: Training History of Hybrid ResNet50 + Swin Transformer Model

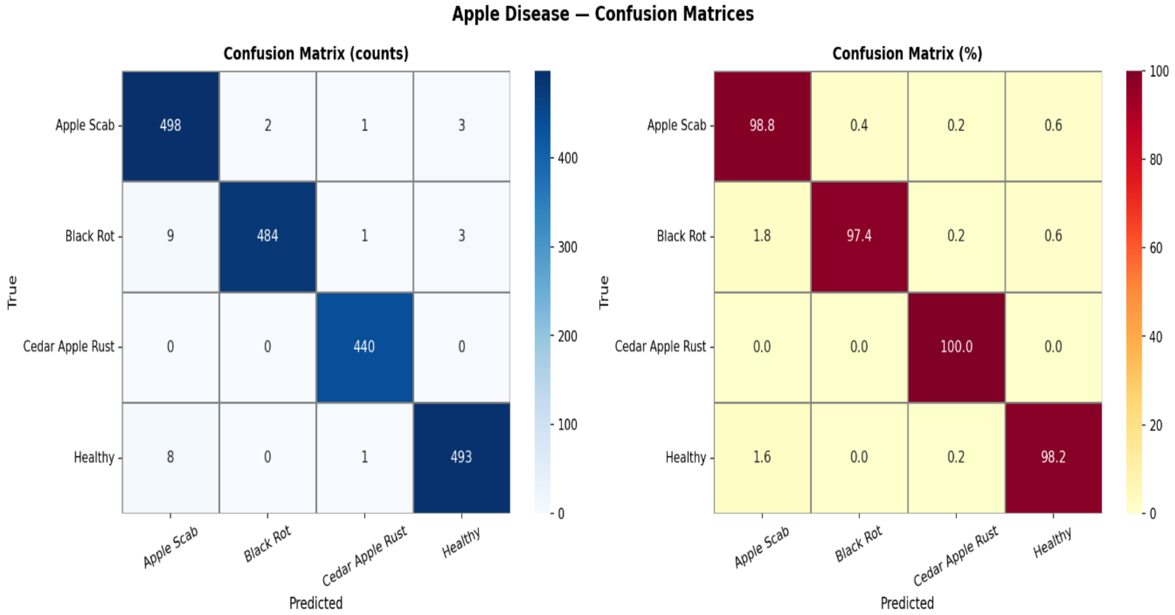
a. Loss Curve Analysis

The left side of the image shows the loss values for both the training and validation datasets, along with the respective training curves for each example. The training loss curves are shown, but it is clear that there is significant initial overfitting, as indicated by the loss starting around 1.45. Initially, the training loss was at a similar level and decreased over the epochs, ending with low values that suggest the hybrid architecture extracted better features. The validation loss similarly declined from approximately 1.05 to below 0.50 by the end of the epochs. Lowering the loss values during training and validation for the apple leaf figure will promote stable model convergence and enhance the learning of specific disease-related characteristics. Most of the times, the value of the loss on the validation set was notably lower than that on the training set. This suggests that the regularization techniques used—such as dropout, mix up augmentation, label smoothing, weighted random sampling, and Bayesian hyper parameter tuning—were effective in minimizing overfitting and enhancing the model's ability to generalize. The consistent trend of convergence without any sudden variance reasserts the reliability of the procedure and effectiveness of the augmented CNN–Transformer design.

b. Accuracy Curve Analysis

On the right side of the window, there are graphs showing accuracy over time for both the training and validation datasets. The starting point for validation accuracy was approximately 65%. At the same point in training, the accuracy hovered around 38%. In the subsequent epochs, there was a consistent and notable rise in training and validation accuracy following the addition of disease-specific features. As the training progressed through several epochs, the accelerated process resulted in validation accuracy exceeding 90%, demonstrating the quick feature development facilitated by the hybrid model design. During the training process, the model's performance steadily improved, and in the final epochs, it nearly achieved a validation accuracy of approximately 98 to 99%. The green dash line of the graph defines a 99% target accuracy. At a peak validation accuracy of 98.56%, despite significant variations in apple leaf samples, the model was nearing the maximum possible validation performance, leaving little room for further improvement. The connection between validation accuracy and training analysis shows that validation accuracy consistently exceeds the training analysis slightly, while the training curve gradually nears an 80% threshold. Although neural networks are intricate, the conflicting observations indicate that the tuning and procedural application are correctly executed, safeguarding against over fitting issues.

C. Confusion Matrix Analysis of Hybrid ResNet50 + Swin Transformer Model



	Apple Scab	Black Rot	Cedar Apple Rust	Healthy
Apple Scab	498	2	1	3
Black Rot	9	484	1	3
Cedar Apple Rust	0	0	440	0
Healthy	8	0	1	493

	Apple Scab	Black Rot	Cedar Apple Rust	Healthy
Apple Scab	98.8	0.4	0.2	0.6
Black Rot	1.8	97.4	0.2	0.6
Cedar Apple Rust	0.0	0.0	100.0	0.0
Healthy	1.6	0.0	0.2	98.2

Figure 3: Confusion Matrix Analysis of Hybrid ResNet50 + Swin Transformer Model

Figure 3 displays the confusion matrix obtained from analysing the Hybrid ResNet50 + Swin Transformer model used to classify apple leaf diseases. This figure consists of two matrices: on the right, it shows the classification accuracy percentages for each disease category, while on the left, it presents the precise number of predictions made for each class. The main focus of the confusion matrix in this research is to assess the model’s ability to predict correctly within classes, as well as to analyse the distribution of correctly and incorrectly identified samples among different apple leaf disease types including Apple Scab, Black Rot, Cedar Apple Rust, and healthy leaves.

a. Confusion Matrix (Counts) Analysis

A breakdown of the exact number of samples included in each category can be seen in the left confusion matrix. The samples that were grouped accurately can be visualized as the items along the diagonal of the matrix in question, whereas the off-diagonal elements illustrate the samples that were misallocated. When 504 samples were put through the classification system, 498 pictures under the category of Apple Scab were correctly identified. Several samples were wrongly classified for Black Rot, Cedar Apple Rust, and Healthy. Describing the lesser number of cases, this implies that the architecture was successful in capturing distinguishing lesion and texture characteristics peculiar to Apple Scab disease. While some samples contained leaves identified as Apple Scab, Cedar Apple Rust, or Healthy, the number of leaves correctly estimated in the Black Rot category is 484. The limited amount of misclassifications proves the effectiveness of the hybrid structure especially for distinction of disease features. As example, the Cedar Apple Rust category showcased a degree of 100% implementation whereby all 440 samples were correctly recognized with no false annotations. This evident coalescing ability is due to the ability in adapting to the structure that is offered and enhancing the features added by the CBAM attention mechanism. Also similarly, the Healthy leaf category depicted very correct classification accuracy in which out of the total 502 images, 493 samples were correctly classified. A very small fraction of the healthy images were depicted as a confused or as diseased leaves. Hence all diseases in the confusion matrix indicate that, apart from the very few misclassifications regarding healthy leaves, the hybrid CNN–Transformer design managed excellent class-wise discrimination capabilities with very low misclassification rates across diseases.

b. Confusion Matrix (%) Analysis

Based on the classification affect, for each given disease, the percentage categorization is given in the confusion matrix on the right. The percentages are most useful for inferring class specific prediction and model performance. For sickness which is in the scope of Apple scab the rate is quite high and is very reliable in a sense that it reaches in 98.8% of classification. In the samples which were shown to be Apple Scab the mistakes committed were of very small percentage. Sequential Tree class has also a high classification accuracy of about 97.4% shape which even is likely to add to the good discrimination effect of the network. The Cedar Apple Rust is 100% reliable within the network system because it can be classified and separated from the other illnesses. Transformer-based global contextual learning lacks completion failures that emphasize efficacy. A healthy leafy area was distinguished from diseased samples, with very few misclassification errors, particularly in the earlier category, achieving an accuracy of 98.2%. The confusion matrix shows very low overall misclassification errors in the lighter off-diagonal regions, while the darker diagonal areas indicate high accuracy in classification performance.

D. Comparative Analysis of YOLOv8s and Hybrid ResNet50 + Swin Transformer Model

The comparison between the YOLOv8s and Hybrid ResNet50 + Swin Transformer models focused on their ability to classify and detect diseases in apple leaves. Evaluation accuracy was determined based on criteria including precision, recall, F1-score, mAP, and overall correctness. The findings show that both models performed their roles as expected, with the ResNet50 combined with the Swin Transformer showing the greatest improvement and overall predictive ability.

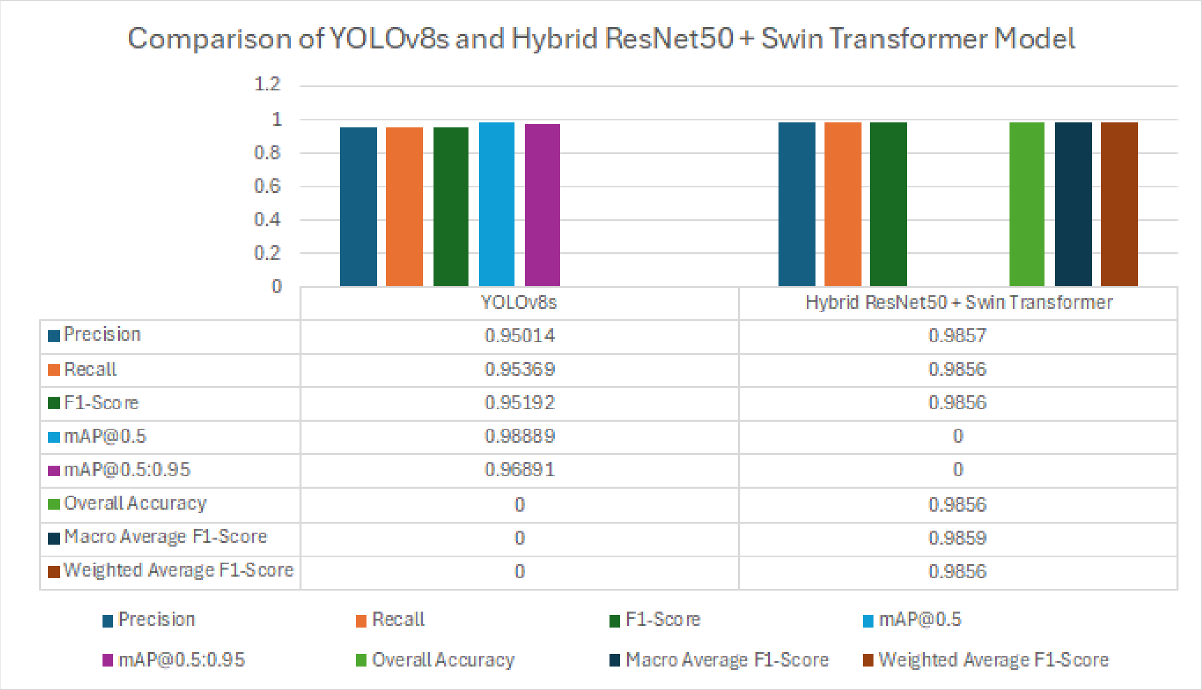
Table 2 presents a comparative analysis between YOLOv8s and the Hybrid ResNet50–Swin Transformer model based on accuracy, precision, recall, F1-score, and disease classification performance for apple leaf disease detection.

Table 2: Comparison of YOLOv8s and Hybrid ResNet50 + Swin Transformer Model

Metric	YOLOv8s	Hybrid ResNet50 + Swin Transformer
Precision	0.95014	0.9857
Recall	0.95369	0.9856
F1-Score	0.95192	0.9856
mAP@0.5	0.98889	—
mAP@0.5:0.95	0.96891	—
Overall Accuracy	—	0.9856
Macro Average F1-Score	—	0.9859
Weighted Average F1-Score	—	0.9856

The precision, recall, and F1-score of the YOLOv8s model are respectively 95.01% 95.36% and 95.19% and they performed well. There were also good mAP@0.5 and mAP@0.5:0.95 values indicating the model performed well in object detection at different IoU levels and localizing diseases precisely. It is therefore hypothesized that YOLOv8s yields good results on an object localization based disease localization task. Unlike this exception, the use of the Hybrid ResNet50 + Swin Transformer model surpassed every other model in the most if not all of the evaluation parameters. Compared to the national benchmarks for YOLOv8s, the classifier outperformed with total accuracy, recall, and F1 scores of 98.56%, 98.56%, and 98.59% respectively, demonstrating strong internal reliability. The most successful accuracy achieved the region Find Cedar Apple Rust and it is the best F1-score on the Benign Class of 99.66%. Also the winnings from all the training images prediction was distributed quite equally. Results with the rest of the classes were also in the very narrow margin from the inter predictor. Lastly the ability of the improved network to gather more precise information globally and locally was demonstrated. It was achieved by the combination of ResNet50 network and Swin Transformer which made an improvement in the ailment classification. The Hybrid ResNet50 and Swin Transformer networks demonstrated improved accuracy of classification over the health of trees. This has most likely been brought about, in favour of the YOLOv8s, its advanced assessment of localization and fast detection. As a result, the combined approach appears to be more suitable for the bulk of work in automated disease control systems and assessing apple leaf disease conditions.

E. Comparison of YOLOv8s and Hybrid ResNet50 + Swin Transformer Model



	YOLOv8s	Hybrid ResNet50 + Swin Transformer
🟦 Precision	0.95014	0.9857
🟧 Recall	0.95369	0.9856
🟩 F1-Score	0.95192	0.9856
🟦 mAP@0.5	0.98889	0
🟪 mAP@0.5:0.95	0.96891	0
🟩 Overall Accuracy	0	0.9856
⬛ Macro Average F1-Score	0	0.9859
🟫 Weighted Average F1-Score	0	0.9856

Figure 4: Comparison of YOLOv8s and Hybrid ResNet50 + Swin Transformer Model

Figure 4 presents the comparative performance analysis of YOLOv8s and the Hybrid ResNet50–Swin Transformer model, showing that the hybrid framework achieved higher accuracy, F1-score, and classification consistency for apple leaf disease detection. An extensive assessment model was used by the researchers to analyse both the YOLOv8s and the Hybrid model. A figure illustrates the comparison of all evaluation metrics related to precision, recall, F1-score, mAP, overall accuracy, and average F1-scores for apple leaf disease classification and detection performed by both models. The YOLOv8s dataset achieved the highest performance in overall image classification tasks. According to the graphical representation, each of the three accuracy scores is greater than 95%. That seems like a fair assumption. Improving the discrimination of different diseases, the scanner achieved better object detection results with increased mAP@0.5 and mAP@0.5:0.95 values, surpassing earlier restrictions. It is generally accepted that YOLOv8s has successfully identified infections in disease detection tasks where precise localization of the affected area is required. In comparison to YOLOv8s, the Hybrid ResNet50 and Swin Transformer duo shows better results for classification, with the Swin Transformer being more effective. The results included high precision and recall rates, F1-scores, and classification accuracy all between 99% and 100%, along with a favorable weighted average F1-score. The existing Hybrid classification model demonstrated consistent performance across all tests, outperforming others, as illustrated in the plot. The assessment demonstrated that the combined Hybrid ResNet50 and Swin Transformer could incorporate diverse features influencing both local and overall appearance, thereby enhancing classification accuracy and prediction effectiveness. In terms of classification performance, the hybrid machine was superior overall, making it more effective for precise diagnosis of apple leaf ailments. Although YOLOv8s can handle detailed detection, it performed less effectively in overall classification compared to the hybrid model, making the latter the better choice for classifying apple leaf diseases.

V. CONCLUSION AND FUTURE WORK

In this research, a pre-trained YOLOv8 model with a Hybrid ResNet50 and Swin Transformer framework was used for automated apple leaf disease spotting in smart agriculture settings. The proposed pipeline was aimed at four classes of apple leaves, meaning Cedar Apple Rust, Apple Scab, Black Rot and Healthy leaves. YOLOv8 showed strong ability in object detection, also good localization, with high accuracy, precision, and mAP scores for locating the diseased regions on apple leaves. The detector managed to capture disease symptoms even when IoU thresholds were changed, so the localization and the classification stayed consistent. After that, the combination of ResNet50 and Swin Transformer improved classification capability a lot, because it merges small-scale feature extraction and broader contextual learning. With this hybrid setup, the system reached 98.56% classification accuracy, and it also reported strong macro and weighted F1-scores. This suggests the predictions were stable, and misclassification between disease categories stayed low. When compared with conventional YOLOv8 benchmarks, the hybrid model gave better feature representation, and stronger disease recognition overall. Overall, the results suggest that Transformer-based hybrid deep learning approaches can meaningfully improve automated plant disease detection and precision farming systems. The proposed framework enables accurate, efficient, and real-time disease surveillance for sustainable smart agriculture. Looking ahead, the same framework can be pushed toward multi-crop disease detection, bigger agricultural datasets, IoT-enabled smart farming systems, drone-based monitoring, and lightweight edge-computing solutions for real-time deployment in the field.

REFERENCES

[1] Raj, Anusha, et al. "YOLO-ODD: an improved YOLOv8s model for onion foliar disease detection." Frontiers in Plant Science 16 (2025): 1551794.
[2] Bao, Weihao, and Fuquan Zhang. "Apple Pest and Disease Detection Network with Partial Multi-Scale Feature Extraction and Efficient Hierarchical Feature Fusion." Agronomy 15.5 (2025): 1043.
[3] Zarboubi, Mohamed, et al. "Enhancing Integrated Pest Management with IoT and YOLO-Evo: A Smart, Low-Cost Monitoring System for Sustainable Apple Farming." Results in Engineering (2025): 108850.
[4] Li, Xiaojuan, et al. "HEFM-YOLO: a lightweight model for apple leaf disease detection in real-world environments." International Conference on Image, Signal Processing, and Machine Learning (ISPML 2025). Vol. 14058. SPIE, 2026.
[5] Erkamim, Moh, Muhammad Zidni Subarkah, and R. Soelistijono. "The Analysis of Architectural YOLOv5 Convolutional Neural Networks for Detecting Apple Leaf Diseases." Journal of Applied Agricultural Science and Technology 9.1 (2025): 40-52.
[6] Zhang, Silu, et al. "YOLO-ACT: An adaptive cross-layer integration method for apple leaf disease detection." Frontiers in plant science 15 (2024): 1451078.

[7] Li, Tong, Liyuan Zhang, and Jianchu Lin. "Precision agriculture with YOLO-Leaf: advanced methods for detecting apple leaf diseases." Frontiers in plant science 15 (2024): 1452502.

[8] Yan, Chunman, and Kangyi Yang. "FSM-YOLO: Apple leaf disease detection network based on adaptive feature capture and spatial context awareness." Digital Signal Processing 155 (2024): 104770.

[9] Jie, Y. U. A. N., et al. "Apple leaf disease detection method based on improved YOLO v7." Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery 55.11 (2024).

[10] Gomez, Daniela, et al. "Advancing common bean (Phaseolus vulgaris L.) disease detection with YOLO driven deep learning to enhance agricultural AI." Scientific Reports 14.1 (2024): 15596.

[11] Yang, Ruotong, et al. "CA-YOLOv5: A YOLO model for apple detection in the natural environment." Systems science & control engineering 12.1 (2024): 2278905.

[12] Reim, Stefanie, et al. "YOLO-Based phenotyping of apple blotch disease (diplocarpon coronariae) in genetic resources after artificial inoculation." Agronomy 14.5 (2024): 1042.

[13] Boudaa, Boudjemaa, et al. "Advancing plant diseases detection with pre-trained yolo models." 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE, 2024.

[14] Sangaiah, Arun Kumar, et al. "UAV T-YOLO-rice: An enhanced tiny YOLO networks for rice leaves diseases detection in paddy agronomy." IEEE transactions on network science and engineering 11.6 (2024): 5201-5216.

[15] Zhou, Pengcheng, et al. "A lightweight apple fruit instance segmentation network: YOLO-AppleSeg." Proceedings of the International Conference on Computer Vision and Deep Learning. 2024.

[16] Xiao, Bingjie, Minh Nguyen, and Wei Qi Yan. "Apple ripeness identification from digital images using transformers." Multimedia Tools and Applications 83.3 (2024): 7811-7825.

[17] Sapkota, Ranjan, Zhichao Meng, and Manoj Karkee. "Synthetic meets authentic: Leveraging llm generated datasets for yolo11 and yolov10-based apple detection through machine vision sensors." Smart Agricultural Technology 9 (2024): 100614.

[18] Anitha, G., et al. "Pest detection and identification in rice crops using Yolo V3 convolutional neural network." 2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0. IEEE, 2024.

[19] Hu, Haipeng, et al. "BHI-YOLO: A lightweight instance segmentation model for strawberry diseases." Applied Sciences 14.21 (2024): 9819.

[20] BalaChandralekha, S., and J. Thangakumar. "Deep learning-based detection of fungal diseases in apple plants using yolov8 algorithm." 2024 International Conference on

Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024.

[21] Zhao, Yun, et al. "Apeiou integration for enhanced yolov7: Achieving efficient plant disease detection." Agriculture 14.6 (2024): 820.

[22] Saputra, Muhammad Andryan Wahyu, et al. "Implementation of YOLO in Cabbage Plant Disease Detection for Smart and Sustainable Agriculture." Brilliance: Research of Artificial Intelligence 4.2 (2024): 798-804.

[23] Mao, DianHui, et al. "Using filter pruning-based deep learning algorithm for the real-time fruit freshness detection with edge processors." Journal of Food Measurement & Characterization 18.2 (2024): 1574-1591.

[24] Dhilleswararao, Pudi & Boppu, Srinivas & Manikandan, M. & Cenkeramaddi, Linga Reddy. (2022). Efficient Hardware Architectures for Accelerating Deep Neural Networks: Survey. IEEE Access. PP. 1-1. 10.1109/ACCESS.2022.3229767.

[25] Kacira, Murat & Ling, Peter. (2001). Design and development of an automated and non-contact sensing system for continuous monitoring of plant health and growth. Transactions of the ASAE. American Society of Agricultural Engineers. 44. 989-96. 10.13031/2013.6231.