

A Hybrid YOLO Backbone–Vision Transformer Framework for Multiclass Mango Leaf Disease Classification

M. Radhika^{a*}, D. Loganayaki^b, V. Priyadharshini^c, J. Mrinalini^d, P. Sivakumar^e, R. Gomalavalli^f, P.G.Gopinath^g

^aDepartment of Electronics and Communication Engineering, R.M.K.Engineering College, Chennai, Tamil Nadu, India.

E-mail: mradhika2211@gmail.com

^bDepartment of Computer Science and Engineering, St.Joseph's College of Engineering, Chennai, Tamil Nadu, India.

E-mail: loganayakid@gmail.com

^cDepartment of Electronics and Communication Engineering, R.M.K.Engineering College, Chennai, Tamil Nadu, India.

E-mail: priya2006inbox@gmail.com

^dDepartment of Electronics and Communication Engineering, R.M.K.Engineering College, Chennai, Tamil Nadu, India.

E-mail: jmrinalini51@gmail.com

^eDepartment of Electronics and Communication Engineering, Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India. E-mail: sivakumar.poruran@gmail.com

^fDepartment of Electronics and Communication Engineering, R.M.K.Engineering College, Chennai, Tamil Nadu, India.

E-mail: gomalavalli@gmail.com

^gDepartment of Electronics and Communication Engineering, R.M.K.Engineering College, Chennai, Tamil Nadu, India.

E-mail: ppg.ece@rmkec.ac.in

Abstract

It is crucial to have the correct diagnosis and timely detection of mango leaf diseases to minimize crop losses and ensure precision agriculture. Handcrafted features or convolutional neural networks (CNNs) have been used in traditional disease classification methods to capture local disease structure, but these traditional methods may not be able to capture the long-range spatial relationships. To tackle this problem, this research put forward a hybrid YOLO Backbone Vision Transformer (YOLO–ViT) for automatic multiclass mango leaf disease classification. The proposed architecture is a fusion of two different architectures: hierarchical feature extraction from a YOLO-based backbone and global contextual representation capability of a Vision Transformer. The framework is used to assess the extent of the following five mango leaf categories: Anthracnose, Golmichi, Healthy, Powdery Mildew and Turning Brown. The confusion matrix shows that 69 out of 75 test samples were correctly classified, resulting in an overall accuracy of 92.0%. The class-wise results suggest that there is complete recognition of Powdery Mildew and Turning Brown samples in the set of tested samples, and major class-wise misclassification is observed between the classes of Anthracnose and Golmichi. The training and validation loss curves show fast initial convergence and relatively stable training behaviour, but there is some variation in the validation loss which requires further generalization analysis. The results reveal the capability of using hierarchical representation from YOLO along with global attention from transformer for mango leaf disease classification and can serve as a foundation for the development of efficient computer vision based disease monitoring systems for precision agriculture.

Keywords: Mango leaf disease classification, YOLO, Vision Transformer, deep learning, computer vision, plant disease detection, precision agriculture.

1. Introduction

Mangifera indica L. (Mango) is one of the most economically significant fruit crops, and maintaining healthy leaves is an essential aspect of sustaining plant growth and productivity. In mango diseases, changes in leaf colour, texture and appearance can be seen, which can be an appropriate source of information for automated disease assessment using leaf images. The visual symptoms may differ from stage to stage, severity to severity, leaf to leaf, and image acquisition to image acquisition, which makes it a difficult computer vision problem to continually recognize the symptoms of the disease. The recent advancements in Artificial Intelligence (AI) have ushered in a new era in the realm of agricultural decision-making support systems that rely on images. The breakthroughs in Artificial Intelligence (AI) have paved new avenues to create image-based agricultural decision-making support tools. The application can be well achieved by deep learning, because it is possible to learn discriminative representations from image data itself, thus reducing dependence on manually designed visual descriptors. Additionally, current computer vision systems provide the means to detect features at multiple spatial resolutions because disease symptoms may manifest as a small patch or a large area across a leaf. But classification is not the only factor in the practical diagnosis of diseases. A helpful system must be able to distinguish between a number of categories of disease and must act uniformly in the classes defined as being visually similar. It should also provide a working architecture based on a computational perspective that may be leveraged for automated agricultural monitoring. These requirements motivate the study of architectures that can take advantage of multiple visual representations instead of relying on a single representation or learning mechanism. In this

study, a multiclass mango leaf disease classification network based on YOLO Backbone–Vision Transformer (YOLO–ViT) is designed. A hybrid YOLO Backbone–Vision Transformer is developed for mango leaf disease classification. The proposed framework is hierarchically structured and has global contextual modelling in a single classification process. The classification behaviour is analysed with the help of class-wise predictions and convergence characteristics, which can give insights into the strengths and weaknesses of the proposed model. The experimental design is used to lay the groundwork for an automatic computer vision-based mango disease monitoring system that can be used in precision agriculture applications. The rest of this paper is organized as follows: The related literature and research gap are presented in Section II and Section III. In Section IV, explains the proposed YOLO–ViT architecture and the experimental results and a detailed performance analysis are given in Section V. The conclusion and suggestions for future research are presented in Sections VI and VII.

II. Literature Survey

Traditional image descriptors and classifiers have slowly been replaced by deep neural networks and attention-based architectures for automated mango leaf disease identification. Manual diagnosis is based on the skill of the practitioners and may be time-consuming if a large plantation area is to be visited. Thus, image-processing and machine-learning techniques have been investigated to get faster and consistent disease detection. But early strategies were very feature-dependent, very image-dependent, disease severity-dependent, and environmental condition-dependent. Srunitha and Bharathi [1] suggested an image processing-based method for identifying unhealthy parts of mango leaves by applying the k-means segmentation algorithm to extract the features and then using the multiclass SVM algorithm for the classification. They showed the feasibility of separating unhealthy areas prior to the classification of the disease, but their need for pre-processing and manually selected representations can limit their adaptability to more complex disease patterns. Chouhan et al., then followed by the development of a web-based anthracnose segmentation system using a radial basis function neural network [2]. It was a first step to transfer traditional image processing to a neural network-based approach for automated anthracnose symptom segmentation in real mango-leaf images. Arivazhagan and Ligi proposed a CNN-based system to identify mango leaf diseases [3]. Their work demonstrated the reduction of dependency on handcrafted descriptors using convolutional feature extraction. Singh et al. also investigated the Multi-Layer CNN for classification of anthracnose and showed that hierarchical deep representations were effective in the diagnosis of the disease [4]. The initial approaches were mostly disease specific or on very limited amounts of data, however, multiclass classification with different visual conditions remained challenging. Mia et al. presented a neural network ensemble and SVM to recognize the mango leaf disease [5]. Overall the reported mean accuracy was about 80% clearly showcasing the possibility of machine learning-based disease recognition and highlighting the challenge of disease class discrimination. Gulvani and Patil applied a deep learning and transfer learning approach to detect mango disease in images [6]. Their experiments had an accuracy of around 91% for mango leaves, with ResNet18 and ResNet34 having an accuracy of 91% and 90.88%, respectively, as reported on the validation settings. While the results show the advantage of pretrained CNN representations, it can differ in other dataset partitions and architectures. Pham et al. [7] studied early classification of mango diseases using feed-forward ANNs with hybrid metaheuristic feature selection. They then pre-processed the images, segmented the images, and selected features before training their evaluated ANN, AlexNet, VGG16, and ResNet50 methods, and obtained 89.41%, 78.64%, 79.92%, and 84.88% accuracy, respectively. This shows that Feature Engineering is able to outperform the original features on small dataset sizes, but may result in dependencies that can impact scalability and robustness. Transfer learning was also applied to investigate AlexNet for mango and grape leaf disease recognition [8]. The accuracy of the mango classification was reported to be about 89%, with precision, recall and F1-score values of about 0.895, 0.905 and 0.90, respectively. It can be seen from the results that the performance of a CNN-based architecture can differ in various crop domains, and that the mango leaf is still difficult to classify automatically, depending on the characteristics of the leaves. Sharma et al. proposed a CNN-based mango leaf disease detection system using an augmented dataset of healthy and diseased leaves [9]. The model achieved a reported accuracy of 90.36%. The study demonstrated the importance of data augmentation for improving CNN-based recognition and highlighted the need for further improvement for accurate disease diagnosis under field conditions and multiclass settings. Thaseentaj and Ilango [10] studied mango leaf disease classification using deep convolutional neural networks for mango leaves in South India. They reported that their customized deep CNN achieved a test accuracy of 91.43%, while their shorter training configuration gave them a test accuracy of 90.20%, the CNN configuration a test accuracy of 53.54%, and the CNN+GAN configuration a test accuracy of 64.04%. The validation behaviour reported also suggested differences between configurations and indicated possible over-fitting. The results show that more complex models do not necessarily result in better generalization. To assess the performance of the MobileNetV2 on mango leaf disease classification, both Simangunsong and Muhathir trained the model on 1500 images of the three classes [11]. Their model performance was 90% accurate, 91% precise, and

90% recall with an F1 score of 89%. The results suggest that lightweight architectures can be employed in the resource constrained applications, and more complex multiclass disease scenarios require stronger feature representations. The overview of studies analyzed is that the traditional machine learning methods have evolved to CNN based feature learning and transfer learning methods. There are still some problems with the existing methods: hand-crafted features, limited disease categories, and the necessity to optimize the method per disease; lack of modelling of long-range spatial relationships; risk of overfitting. In addition, mango disease classification challenges are also mentioned by Faye et al. [12] because of the scarcity of data, environmental conditions and implementation of deep-learning. In this case, the efficient use of computers is of essential importance for agriculture. Tan and Le have introduced EfficientNet, and shown that scaling a network in terms of depth, width, and input resolution can lead to good performance while avoiding overkill [17]. Another important line is the YOLO based architectures due to their efficient feature extraction and prediction. YOLO-based methods are thus promising to be a good foundation for building effective visual feature extraction systems in agriculture. The literature reviewed shows that there is a gap in the research between the use of conventional CNN-based mango disease classification and architectures that can make use of local and global representations simultaneously as shown in Table 1. The former approaches were often based on handcrafted features or traditional CNNs, while transformer-based models offer ways to capture the global contextual relationships. Hence, YOLO BackboneVision Transformer (YOLO-ViT) is proposed to integrate the hierarchical convolutional feature extraction with the transformer-based global attention. The goal is not only to increase the accuracy of numbers but also to increase the representation of local patterns and spatial relationships pertinent to a particular disease.

Table 1: Comparison of Existing Mango Leaf Disease Classification Methods

Ref.	Approach	Reported Performance	Main Observation
[1]	K-means + Multiclass SVM	Reported as effective	Depends on segmentation and handcrafted representations
[2]	RBF Neural Network	—	Automated anthracnose segmentation
[3]	CNN	—	Early CNN-based mango disease recognition
[4]	Multilayer CNN	—	Disease-specific deep feature learning
[5]	Neural Network Ensemble + SVM	80%	Moderate multiclass recognition
[6]	ResNet Transfer Learning	90.88–91.50%	Transfer learning improves representation
[7]	ANN + Metaheuristic Feature Selection	89.41%	Effective but preprocessing dependent
[8]	AlexNet Transfer Learning	89%	Limited mango-domain performance
[9]	CNN + Data Augmentation	90.36%	Augmentation improves recognition
[10]	Customized Deep CNN	91.43%	Some evidence of validation fluctuation
[11]	MobileNetV2	90%	Lightweight but moderate performance

III. Research GAP

The existing techniques have been known to the limitations in representation, generalisation, and computation, despite the achievements in automated mango leaf disease classification. The handcrafted features in previous machine learning models rely heavily on the experience of the experts, and CNN-based models are mostly used to capture the local spatial information and may not be adequate in modelling the long-range relationship between disease regions. Transfer-learning techniques have resulted in better feature extraction; however, they are still limited by the properties of the pretrained architecture as well as the dataset. In addition to the study of existing methodologies, we have reported that the moderate performance analysis might require some additional preprocessing or feature selection steps. The literature surveyed indicated that there is not much research conducted on combining YOLO with Vision Transformer (ViT) into a single architecture for multiclass mango leaf disease classification, which is able to efficiently

extract features from the leaves based on their hierarchy, in addition to modelling the global context of the whole image. There is, therefore, a need for the development and testing of such a hybrid representation in a systematic manner, particularly when it comes to representational categories that are visually similar (e.g., Anthracnose and Golmichi). This spurs our proposed YOLO Backbone–ViT framework. Table 2 shows the major limitations of the current mango leaf disease classification methods and the proposed research directions to overcome these limitations within the proposed YOLOViT framework.

Table 2: Research Gap and Motivation for the Proposed Framework

Existing limitation	Consequence	Proposed direction
Handcrafted feature dependence	Limited adaptability	Learned hierarchical representations
Disease-specific models	Poor scalability to multiple classes	Multiclass classification
CNN local receptive-field emphasis	Limited global context	ViT self-attention
Large/complex architectures	Deployment difficulty	YOLO-derived efficient backbone
Dataset-specific optimization	Weak generalization	Robust evaluation and class-wise analysis
Similar disease symptoms	Misclassification	Local + global feature fusion
Accuracy-only evaluation	Limited diagnostic insight	Confusion matrix + multiple metrics

IV. Proposed model

The overall architecture of the proposed YOLO Backbone–Vision Transformer framework for multiclass mango leaf disease classification is shown in Figure 1. A hybrid YOLO backbone and Vision Transformer (YOLO – ViT) is proposed for the multiclass mango leaf disease classification in this work. The designed architecture integrates the hierarchical feature extraction ability of a YOLO-like backbone with the global contextual representation ability of Vision Transformer (ViT). The objective is to improve the accuracy of distinguishing between two similar-looking mango leaf diseases while maintaining an efficient feature learning process.

A. Dataset Preparation

The data was taken from five categories of mango leaves— Anthracnose Golmichi, Healthy, Powdery Mildew, and Turning Brown. To ensure uniform size of the input images, all images are first resized to 224 × 224 pixels. Some of the dataset augmentation techniques that were used are rotating images, horizontal flipping, zooming in and out, changing the brightness and scaling the images to increase variation in the data and reduce the issue of overfitting. The dataset is divided into training, validation and testing sets. In the first stage of the proposed approach, a feature extraction backbone network is used to extract the discriminative visual features from the mango leaf images using YOLO. The YOLO backbone is a new backbone network that outputs multi-level feature representations, which are unavailable to traditional CNNs that sequentially perform convolutional operations.

Let the input image be denoted by:

$$I \in \mathbb{R}^{224 \times 224 \times 3} \quad (1)$$

YOLO backbone maps the input image to feature maps:

$$F = f_{\text{YOLO}}(I) \quad (2)$$

Whereas, I = input image, F = Feature representation extracted from text. f_{YOLO} represents the extraction of features from YOLO. These features include disease information, like texture variation, colour changes, lesion structures, edge information, and spot distribution. Feature maps are extracted and then transformed into feature tokens for further processing by a transformer.

B. ViT (Vision Transformer):

To use YOLO backbone for feature maps, they are divided into patches, and token embeddings are created. For patch embedding:

$$Z_0 = [x_p^1 E; x_p^2 E; \dots; x_p^N E] + E_{\text{pos}} \quad (3)$$

Whereas, x_p is the image patch, E represents the embedding matrix, and E_{pos} refers to the positional encoding. The transformer encoder is designed to model long-range relationships between disease regions with multi-head self-attention:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Whereas Q is the Query matrix, K is the Key matrix, V is the Value matrix, and d_k represents a key vector dimension. The self-attention mechanism enables the model to detect a relationship between the symptoms of a diseased area, which are separated from each other.

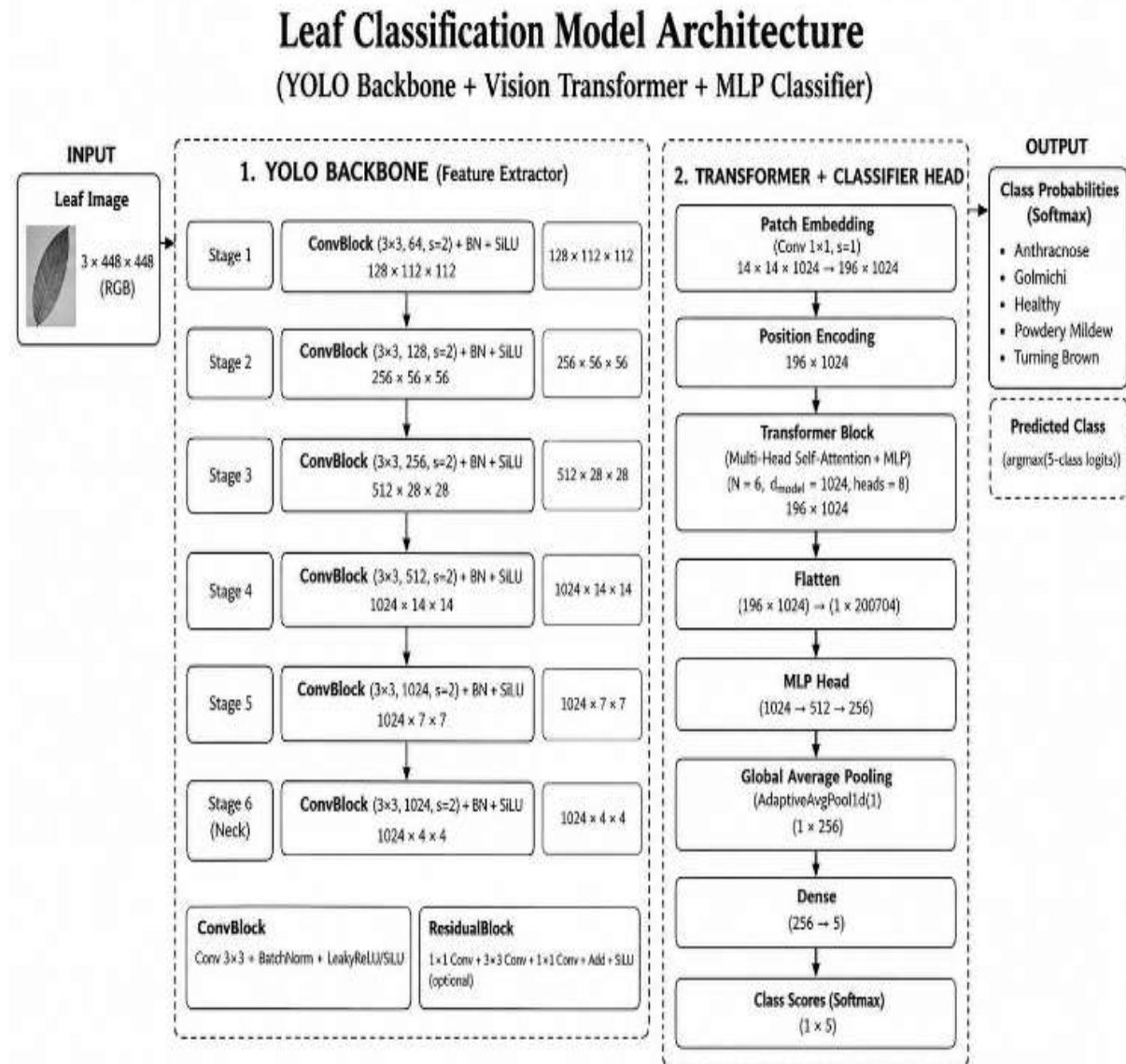


Figure 1: An Architecture of the Proposed Hybrid YOLO Backbone–Vision Transformer Framework for Multiclass Mango Leaf Disease Classification

C. Classification Layer

The output tokens of the transformer encoder are fed through fully connected layers with a SoftMax activation function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^n e^{z_j}} \quad (5)$$

where: $P(y_i)$ represents the class size, i.e., the number of students in each class. Z_i is the output score, n shows the number of classes. The final class will be:

$$\hat{y} = \arg \max(P(y_i)) \quad (6)$$

V. Experimental Result

The results of the experiments are presented and discussed. The proposed YOLO–ViT was evaluated against five categories of mango leaves which includes: Anthracnose, Golmichi, Healthy, Powdery Mildew, and Turning Brown, and evaluated using the following analysis:

Table 3: Class-Wise Correct Predictions of the Proposed YOLO-ViT Model

Actual Class	Correct Predictions	Total Samples
Anthracnose	11	15
Golmichi	14	15
Healthy	14	15
Powdery Mildew	15	15
Turning Brown	15	15

In Table 3, shows the class-wise prediction results of the proposed YOLO-ViT model, where the number of correctly classified samples for each of the mango leaf classes in the test set is indicated. The total number of correctly classified samples is: $11+14+14+15+15=69$. Total test samples are 75. Therefore, the accuracy of classification is: $\text{Accuracy} = \frac{69}{75} \times 100$ $\text{Accuracy} = 92.0\%$ The confusion matrix shows that Powdery Mildew and Turning Brown are well explained. If the samples are healthy, then there is only one misclassification. There are two main confusions: between Anthracnose and Golmichi. Some of the same signs of appearance lead to the misclassification.

Training and Validation Loss Analysis:

The training loss curve is very steep in the initial epochs. The reduction in training loss ranges from 0.95 to 0.05. When the first loss of validation occurs it starts to have a significant impact. After the 20th epoch, small fluctuations can be seen. The curve around epoch 24 has a peak in the validation curve that is temporary. This results in the model stabilising after the next epochs. The decrease of loss values shows that the model can be used to capture disease-specific characteristics.

Performance Evaluation:

The proposed model is capable of achieving an accuracy of 92.0%. The integration of the YOLO hierarchical feature extraction and the global attention layer in Vision Transformer. Improves the representation of features for multiclass disease identification. From the results above, it can be seen that the convolutional feature extraction and the transformer attention are complementary in the disease recognition task. The local lesion structures, the texture and colour differences are extracted by the YOLO backbone, and the Vision Transformer models long-range relationships between the disease regions. The confusion matrix shows that diseases that are visually similar are difficult to manage, especially Anthracnose and Golmichi. This means that there is a need for improvements in the future through larger datasets for the Explainable AI techniques to have efficient attention visualization, hyperparameter optimization, and multi-scale feature fusion. In conclusion, the proposed YOLO-ViT framework is effective in automated mango leaf disease cataloguing, in future it might be used in precision agriculture applications.

Confusion Matrix

The analysis of confusion matrix for Anthracnose shows that 11 out of 15 Anthracnose samples were correctly classified, while 4 samples were classified as other classes. In addition, 59 samples from the other categories were correctly identified as other classes, whereas 1 sample from the other categories was incorrectly predicted as Anthracnose. The results indicate that Anthracnose is one of the more challenging classes for the proposed model, with a higher number of misclassified samples compared with the other disease categories. Figure 2 illustrates the confusion matrix obtained for the Anthracnose class.



Figure 2: Confusion Matrix for Anthracnose Classification Using the Proposed YOLO-ViT Model

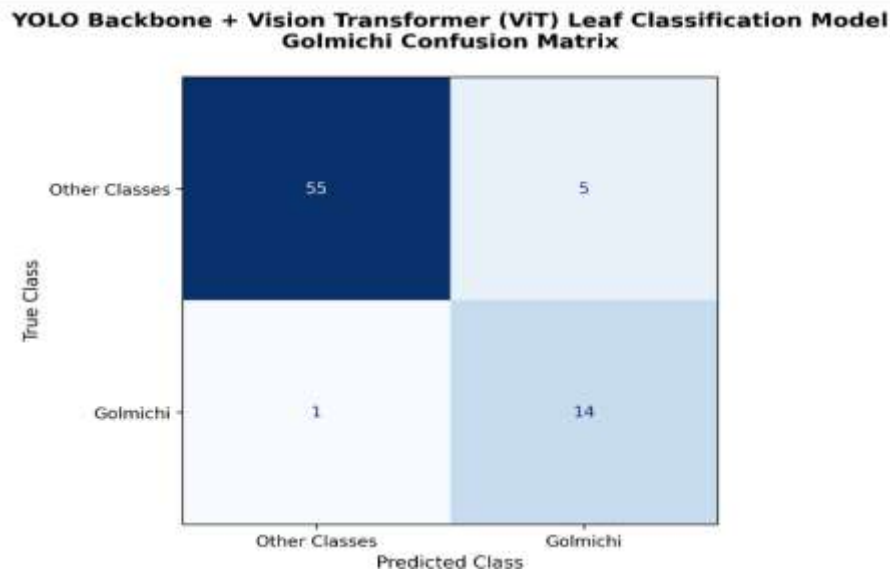


Figure 3: The confusion matrix for Golmichi classification by using the proposed model named YOLO-ViT is presented

For Golmichi 14 samples were correctly classified; 1 sample of Golmichi was classified as another class. Other variables were used as well, but were not as effective. Therefore, in spite of the good recognition of the actual Golmichi samples, the relatively high number of false-positive predictions points towards difficulties in discriminating Golmichi-like visual patterns from patterns of other leaf conditions. The confusion between Anthracnose and Golmichi has been observed and fits with the visual similarity of these categories reported during the experimental analysis. The confusion matrix for the Golmichi class is shown in Figure 3.

Of the samples labelled as the Healthy category, 14 were correctly classified with one sample classified as another. All other categories have none of their samples misclassified as Healthy in the presented matrix, showing that the model is very accurate at differentiating between healthy and diseased leaves. Figure 4 shows the confusion matrix for the Healthy class.

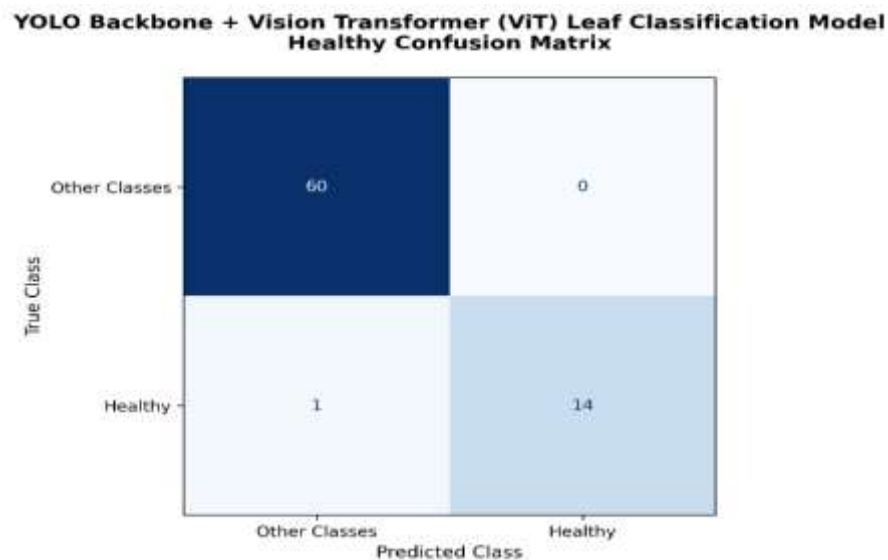


Figure 4: Confusion Matrix for Healthy Leaf Classification Using the Proposed YOLO-ViT Model

The matrix of the Powdery Mildew class shows recognition of all pixels. All 15 samples of Powdery Mildew were correctly classified and no samples from other categories were incorrectly predicted as Powdery Mildew. So, in the test set analysed, the model has performed well with Powdery Mildew with 100% classification accuracy. The confusion matrix is shown in figure 5 for the Powdery Mildew class.

Likewise, the class Turning Brown was completely recognized in the tested set. No Turning Brown sample was misclassified as other categories and all 15 Turning Brown samples were correctly classified. Furthermore, none of the samples of the other classes was wrongly classified as Turning Brown. The Turning Brown category has the highest probability of being recognized by the proposed model, suggesting

it is one of the most consistent categories. The confusion matrix found for the Turning Brown class is shown in Figure 6.

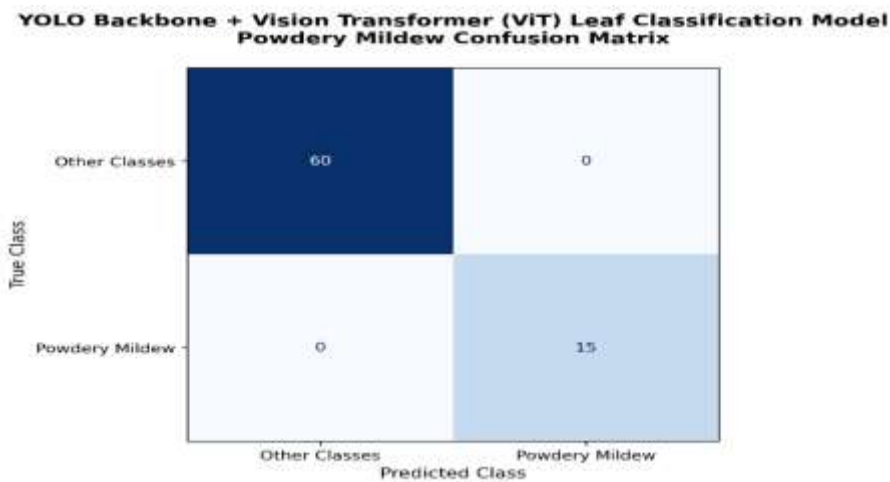


Figure 5: Confusion Matrix for Powdery Mildew Classification Using the Proposed YOLO-ViT Model



Figure 6: Confusion Matrix for Turning Brown Classification Using the Proposed YOLO-ViT Model

The confusion matrix for each class shows that the proposed YOLO-ViT model achieves excellent results for the Powdery Mildew and Turning Brown classes with 100% accuracy on all test samples. The recognition performances of Healthy and Golmichi are also good, obtained 14 classification out of 15 samples for each class. The highest percentage of misclassifications is seen with Anthracnose at 4 of 15 misclassifications. The overall conclusion from the confusion matrix analysis is that the major difficulty is to differentiate between similar images of the disease such as those of Anthracnose and Golmichi. This observation aligns with the motivation of the proposed hybrid architecture that integrates the hierarchical features extracted by the YOLO backbone with the global contextual modeling that is done by the Vision Transformer.

VI. Conclusion

The YOLO Backbone-Vision Transformer (YOLO-ViT) framework has been proposed for the multiclass mango leaf disease classification problem. The proposed approach combines hierarchical feature extraction and global contextual modelling with transformers to classify into five classes: Anthracnose, Golmichi, Healthy, Powdery Mildew, Turning Brown. For the test set used, the model was able to correctly identify 69 out of 75 samples with an accuracy rate of 92.0%. The "confusion matrix" indicates that the classes Powdery Mildew and Turning Brown are fairly easy to recognise, but the classes Anthracnose and Golmichi are still somewhat challenging. The training and validation loss behaviour shows that our proposed hybrid YOLO-ViT model performs well in mango leaf disease classification.

VII. Future Work

Future work will be directed towards testing the proposed framework with larger and geographically more diverse mango-leaf datasets to test its generalisation under varying environmental conditions. In the coming weeks, more improvements will delve into the use of data augmentation, hyperparameter tuning, attention mechanisms for feature fusion, and explainable AI approaches such as Grad-CAM and attention visualization. The ablation studies of YOLO-only, ViT-only and hybrid will be conducted again to perform

further experiments. Parameters of the model, model complexity, training time and inference latency will also be analyzed to make sure it is appropriate for real-time agricultural use. Lastly, there is potential to connect with mobile or edge-computing platforms for conducting practical field-level mango disease monitoring.

References

- [1] K. Sruthi and D. Bharathi, "Mango leaf unhealthy region detection and classification," in *Computational Vision and Bio Inspired Computing*, Cham, Switzerland: Springer, 2018, pp. 422–436, doi: 10.1007/978-3-319-71767-8_35.
- [2] S. S. Chouhan, U. P. Singh, and S. Jain, "Web facilitated anthracnose disease segmentation from the leaf of mango tree using radial basis function (RBF) neural network," *Wireless Pers. Commun.*, vol. 113, pp. 1279–1296, 2020, doi: 10.1007/s11277-020-07279-1.
- [3] S. Arivazhagan and S. V. Ligi, "Mango leaf diseases identification using convolutional neural network," *Int. J. Pure Appl. Math.*, vol. 120, no. 6, pp. 11067–11079, 2018.
- [4] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer convolution neural network for the classification of mango leaves infected by anthracnose disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019, doi: 10.1109/ACCESS.2019.2907383.
- [5] M. R. Mia, S. Roy, S. K. Das, and M. A. Rahman, "Mango leaf disease recognition using neural network and support vector machine," *Iran. J. Comput. Sci.*, vol. 3, no. 3, pp. 185–193, 2020, doi: 10.1007/s42044-020-00057-z.
- [6] S. Gulvani and R. Patil, "Deep learning for image based mango leaf disease detection," *Int. J. Recent Technol. Eng.*, 2019.
- [7] T. Pham, L. V. Tran, and S. V. T. Dao, "Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection," *IEEE Access*, vol. 8, pp. 189960–189973, 2020.
- [8] "Deep learning precision farming: Grapes and mango leaf disease detection by transfer learning," *Global Transitions Proc.*, vol. 2, no. 2, pp. 535–544, 2021, doi: 10.1016/j.gltp.2021.08.002.
- [9] S. Sharma et al., "Mango leaf diseases detection using deep learning," *Int. J. Knowl.-Based Comput. Syst.*, 2022.
- [10] S. Thaseentaj and S. S. Ilango, "Deep convolutional neural networks for South Indian mango leaf disease detection and classification," *Comput. Mater. Contin.*, vol. 77, no. 3, pp. 3593–3618, 2023, doi: 10.32604/cmc.2023.042496.
- [11] R. C. Simangunsong and Muhathir, "Mobilenetv2 analysis in classification diseases on mango leaves," *J. Informatics Telecommun. Eng.*, vol. 8, no. 3Spc, 2025, doi: 10.31289/jite.v8i3Spc.14430.
- [12] D. Faye, I. Diop, and D. Dione, "Mango diseases classification solutions using machine learning or deep learning: A review," *J. Comput. Commun.*, vol. 10, no. 12, pp. 16–28, 2022, doi: 10.4236/jcc.2022.1012002.
- [13] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosyst. Eng.*, vol. 172, pp. 84–91, 2018, doi: 10.1016/j.biosystemseng.2018.05.013.
- [14] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agric.*, vol. 161, pp. 272–279, 2019, doi: 10.1016/j.compag.2018.03.032.
- [15] J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanekhan, "Using deep transfer learning for image-based plant disease identification," *Comput. Electron. Agric.*, vol. 173, Art. no. 105393, 2020, doi: 10.1016/j.compag.2020.105393.
- [16] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2021.
- [17] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019.
- [18] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.