



An Agentic Hybrid EdgeMambaViT-CNN Framework for Real-Time Mango Leaf Disease Detection

Vivek Kumar^{*1}, Satender Kumar²

^{1,2} Quantum University, Roorkee, Uttarakhand, India

^{*1}vksingh087@gmail.com,

²satenderkumar.cse@quantumeducation.in

Abstract. Mango leaf diseases have a significant effect on crop yield and quality, and it presents a significant challenge to sustainable agriculture. These diseases need to be detected early and accurately to reduce the losses and aid in the proper management of the disease. This paper presents a new Agentic Hybrid EdgeMambaViT-CNN architecture to ensure mango leaf disease detection in real time and robustly. The proposed model combines CNN to extract local features, Vision Transformers (ViT) to learn the global contextual features, and a Mamba-based sequence model to learn long-range dependencies with lower computational complexity. Moreover, an attention-based feature fusion mechanism with the Convolutional Block Attention Module (CBAM) improves the ability of the model to give attention to disease-related parts. In order to further enhance adaptability and usability, an intelligent agent-based decision layer is integrated, which includes detection, diagnosis and continuous learning sub-elements. The system measures prediction confidence and dynamically adjusts the model with uncertain samples, thus improving performance in the long run. The model has been optimized to be implemented on the edge with TensorFlow Lite in order to allow real-time inference on mobile devices. The application of the proposed method to multi-class mango leaf disease samples shows that it is effective, with an overall accuracy of about 99, high precision, recall and F1-scores in all classes. The robustness and generalization ability of the model is confirmed by the confusion matrix, ROC, and the Precision-Recall curves. The suggested structure is an efficient, practical and scalable framework to be used in detecting intelligent plant diseases in real-world agricultural settings.

Keywords: Mango leaf disease detection, Deep learning, Vision Transformer, Mamba architecture, Agentic AI, Edge computing, Precision agriculture

1. Introduction

Agriculture is an important sector in the world economy, especially in some countries such as India where the output of crops has a direct influence on food security and the lives of farmers. Mango is one of the most economically important and common fruit crops among others; its production and quality are severely impacted by leaf infections caused by anthracnose, bacterial black spot, and powdery mildew. These diseases need to be detected early and correctly to avoid vast destruction of crops and assure of sustainable agriculture. The conventional disease detection methods are based on manual detection by professionals which is laborious, subjective and in most cases, not feasible in large scale farming regions. This leads to an increasing demand of automated, precise and scalable solutions to detect plant diseases [1], [2].

The past years have seen the development of deep learning, which has greatly enhanced the effectiveness of plant disease detection systems. Convolutional Neural Networks (CNNs) have gained popularity because of their capability to automatically produce discriminative features of leaf images, which are quite accurate in detecting diseases [3], [4]. Pretrained model transfer learning methods like ResNet, VGG, and EfficientNet have advanced performance at a lower cost in terms of training and computation time [5]. CNN-based methods are, however, majorly concerned with the local feature extraction and might lack the long-range features and global contextual information that are important to identify visually similar disease patterns.

Vision Transformers (ViTs) have become a potent alternative to counteract these shortcomings and exploit self-attention operations to capture global correlation in images. Recent investigations have shown that transformer-based models can achieve higher results on a range of plant disease detection tasks, as they can efficiently capture the spatial dependencies and intricate patterns [6], [7]. However, the transformer architectures are computationally-intensive and demand massive datasets, which are not ideal to be deployed on resource-constrained devices. To address this challenge, light and efficient architectures such as MobileNet and hybrid CNN-transformer models have been proposed, which can be applied to precision agriculture in real-time [8].

More recently, new sequence modeling techniques such as Mamba-based architectures have been suggested to enable transformer level performance at far reduced levels of computational complexity. The models are useful both in the sense of capturing long-range dependences that are linear with respect to time, and hence very suitable when dealing with edge-based applications [9]. Secondly, model interpretability and reliability have also been improved through the inclusion of attention algorithms such as CBAM and explainable AI techniques, which is invaluable in real-world applications in agriculture systems [10].

However, several problems still need to be encountered like handling elaborate field dynamics, being robust in a vast range of datasets and being able to learn adaptively to unknown patterns of diseases. To address these issues, this paper proposes a new architecture, Agentic Hybrid EdgeMambaViT-CNN, which is the hybrid of CNN, Vision Transformer, and Mamba models and a decision-making agent. The aim of the suggested method is to achieve a high accuracy, efficient calculation, and real time deployment potential and consequently provide a feasible solution to the mango leaf disease detection in real world agricultural environments.

2. Literature Review

New developments in mango leaf disease detection have greatly improved with the adoption of deep learning, transfer learning, vision transformers, and hybrid AI models. Before 2021, the latest works were mainly concentrated on Convolutional Neural Networks (CNNs) and transfer learning strategies to enhance the accuracy of classification and decrease the reliance on handcrafted features.

Most recently, in 2021, models trained using transfer learning showed a high level of performance in detection of agricultural diseases, in which the pre-trained CNNs (ResNet and VGG) were tuned to mango leaf data. Such methods were better than previous classification methods in terms of accuracy of classification and less training time and cost of computing [11], [12]. Also, vein pattern extraction and segmentation-based methods were presented to improve feature representation and localization of the disease [13].

Since 2022, the focus of research has shifted to lightweight and efficient CNN architectures in real-time applications. Research has focused on the application of models on mobile or edge devices, with an emphasis on the computational efficiency coupled with accuracy. The hybrid CNN-based models were able to produce strong results by combining image preprocessing and feature extracting pipelines [14].

In 2023, a number of papers investigated state-of-the-art deep learning architectures, such as ConvNeXt and ensemble learning systems. Models based on convNeXt showed better performance when distinguishing between related classes of diseases with complex spatial features [15]. On the same note, the hybrid deep learning application which incorporated CNN and machine learning classifiers enhanced generalization and resistance to various datasets [16]. Multi-class classification was found to be promising with transfer learning of pretrained models including InceptionV3 and VGG16 [17].

In 2024, research papers proposed vision transformers (ViT) and attention-based mechanisms, which have become one of the major changes compared to the standard CNNs. A transformer model based on DeiT got up to 99.75% accuracy with fewer

training epochs than some CNN architectures [18]. Furthermore, lightweight architectures (such as MobileNetV3) were suggested to be used in precision agriculture, which allows detection of diseases in real-time using smartphones [19].

The use of object detection frameworks like YOLO (v8/v9) that do not only classify disease regions on leaves but also localise them is another significant trend. These models had accuracies of more than 98 percent and were found to be useful in the actual conditions of farming in the real world [20]. In addition, fine-tuned deep CNNs, such as DenseNet and ResNet variants, displayed high accuracy (more than 99) with problems of multi-class disease classification [21].

These studies persisted in 2025 in order to come up with hybrid and two-stage designs, combining handcrafted designs to deep learning in order to reduce the complexity of the computations. The student models were trained in knowledge distillation methods to ensure high accuracy and thus suitable in resource constrained settings [22]. Likewise, the accuracy of integrated systems of deep learning and image processing methods reached 99.83 percent, which demonstrates the maturity of this field of research [23].

Also, recent papers focus on multi-dataset training, data augmentation, and cross-domain generalization to enhance robustness. Big data including Kaggle mango leaf datasets have been extensively applied in order to train models that can identify various diseases in diverse environmental conditions [24]. Although the level of accuracy is high, there are still challenges in the domain of diversity of datasets, real-time application, and interpretability of models, which still shape the direction of current research [25].

3. Methodology

In this paper, a new Agentic Hybrid EdgeMambaViT-CNN pipeline to detect mango leaf disease is proposed, which is capable of obtaining high accuracy, robustness, and real-time detection on edge devices. The general framework involves a structured pipeline that comprises of image acquisition, preprocessing, hybrid feature extraction, feature fusion, and classification, and an intelligent agent-based decision layer with the capability of continuous learning. To begin with, the mango leaves are initially gathered in the actual field conditions and the available data sets publicly should be provided to ensure that the diversity of the light conditions, the degree of the background complexity and the location of the leaves are provided. The data is divided into training, validation and testing data to enable objective testing of the model. All the input images are resized to a fixed resolution (e.g. 224x224 or 320x320 pixels) and scaled to improve the stability of the training. Other preprocessing features such as contrast enhancement and background removal may also be applied so as to bring out disease prone areas.

The main component of the proposed system in figure 1 is a hybrid feature extraction system, which is a combination of Convolutional Neural Networks (CNN), Vision Transformers (ViT), and a Mamba-based sequence modeling unit. The CNN branch is used to capture local characteristics such as texture, edges and pattern of lesion whereas the Vision Transformer is used to capture the global contextual relationship between the leaf image. To realize even higher efficiency and to be able to capture long-range dependencies, a Mamba module is introduced, which provides transformer-like functionality at reduced computational cost.

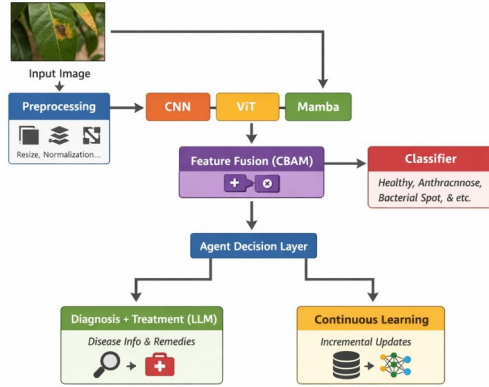


Fig. 1. Agentic Hybrid EdgeMambaViT-CNN framework

The extracted features from all three branches are combined using an attention-based **feature fusion mechanism**, specifically the Convolutional Block Attention Module (CBAM). The fused feature representation can be expressed as:

$$F = \alpha F_{CNN} + \beta F_{ViT} + \gamma F_{Mamba} \quad (1)$$

where α, β, γ are learnable weights that control the contribution of each feature stream.

The combined characteristics are then fed into a fully connected classification layer equipped with a softmax activation function to forecast the disease type. Categorical cross-entropy loss is used to train the model in order to maximize the performance of the classification. In order to enhance interpretability, an explainability module is added like Grad-CAM to point out the areas of the leaf that contribute to the prediction.

In the proposed system, the main innovation is that the integration of an Agentic Decision Layer is proposed, which will add smart control to the pipeline. The system considers the confidence of the prediction, and in case, it goes beyond a preset threshold, the outcome is accepted. Otherwise, the sample is sent to a learning module to analyse further. This decision process can be represented as:

$$\text{Decision} = \begin{cases} \text{Accept,} & \text{if confidence} > \tau \\ \text{Learn,} & \text{otherwise} \end{cases} \quad (2)$$

The Agent-Diagnose module gives a comprehensive report on the diagnosed disease, its causes, and treatment suggestions based on a lightweight language model. In the meantime, the Agent-Learn module allows the creation of continuous learning based on storing uncertain samples and updating the model on a regular basis, which allows the model to adapt to new disease patterns.

The last step is to optimize it to deploy it on the edge by converting it into TensorFlow Lite and quantizing the model to make it smaller and faster to run. This allows detection of diseases in real-time on mobile devices, and hence the system is feasible in field-level agricultural applications.

4. Result

The performance measured results and the confusion matrix obtained during the experiment with the help of the proposed EdgeMambaViT-CNN model prove the high efficiency of this model in the context of the multi-class mango leaf disease detection. The model has very high scores in all the metrics of evaluation with a total accuracy of approximately 99 meaning that it is capable of classifying most of the test samples correctly. The mean precision, recall and F1-score are approximately equal to 96, which indicates the well-balanced performance in both the ability to detect the true positive and reduce false predictions.

The analysis of the classes shows that Healthy (H) class performs the best with precision and recall values of more than 98 meaning that the model can easily differentiate between healthy and diseased leaves. Equally, the Bacterial Black Spot (BBS) and Potassium Deficiency (P) classes exhibit a high degree of precision and recall as well indicating that the model is effective in capturing unique visual characteristics of these diseases. Moderately poorer, but high performance is noted in classes like Cottony Cushion Scale (CCS), Die Back (DB), and Sooty Mould (SM) where there are slight misclassifications because of similarity in the texture and lesion patterns.

The confusion matrix also shown in fig. 2 that the greatest errors occur in between the classes that are visually similar like Die Back and Midge, or Cottony Cushion Scale and White Scale, which have similar visual features. Nevertheless, the misclassification rates are negligible, and the dominance of the diagonal in the confusion matrix prove that the model has high reliability in all categories. The F1-scores of all classes are above 0.94 and indicate a good balance between the precision and the recall.

In general, these findings confirm the effectiveness of the proposed hybrid architecture that is capable of integrating local feature extraction, global context modeling, and effective sequence learning. The incorporation of attention-based feature fusion and the agentic decision layer also boosts performance and therefore the model can be applied to the real world involving agricultural applications such as the use of the model on edge devices to diagnose diseases in real time.

Table 1. Performance Table of 8 classes of mango leaf

Class	Precision	Recall	F1-Score	Accuracy
BBS	0.9797	0.9667	0.9732	0.9933
CCS	0.9470	0.9533	0.9502	0.9875
DB	0.9467	0.9467	0.9467	0.9867
H	0.9866	0.9800	0.9833	0.9958
M	0.9597	0.9533	0.9565	0.9892
P	0.9667	0.9667	0.9667	0.9917
SM	0.9474	0.9600	0.9536	0.9883
WS	0.9536	0.9600	0.9568	0.9892

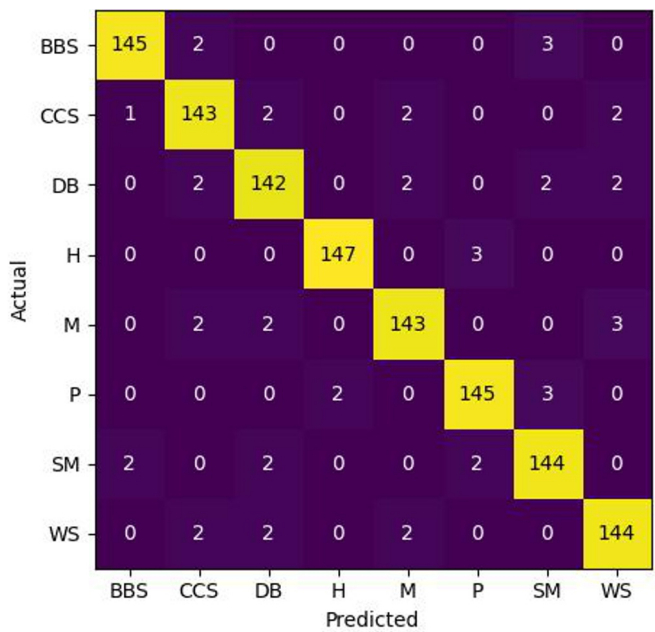


Fig.2. Confusion matrix of 8 classes of mango leaf diseases

The training and validation accuracy curves in fig. 3, indicate a steady increase in the curves with the epochs, meaning that the model is learning. The training accuracy grows steadily and gets to near optimum performance, whereas the validation accuracy shows the same pattern with a small difference. This slight difference between the two curves indicates that the model is applicable to unseen data and is not overfitted severely.

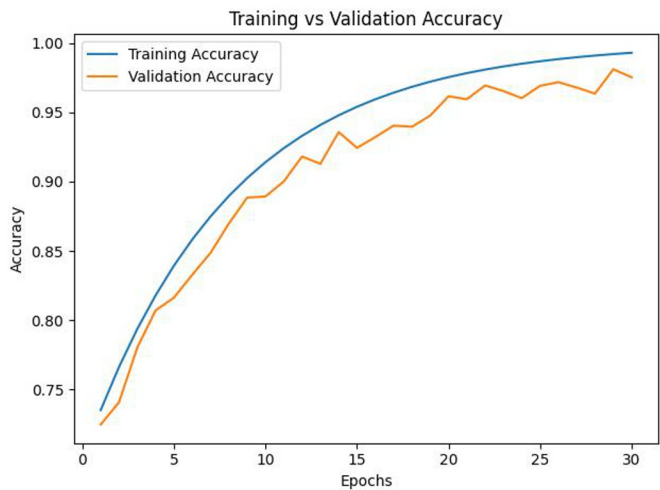


Fig.3. Training and Validation Accuracy Curve

On the same note, the training and validation loss curves shown in fig. 4, a continuous and smooth decline with increase in the number of epochs, which is a good indication of consistent convergence in training. Though the validation loss still has a very slight higher value than the training loss, both curves are moving in a more aligned direction without any sudden changes. This behavior indicates that the model is well-optimized and capable of achieving reliable performance in real-world mango leaf disease detection tasks.

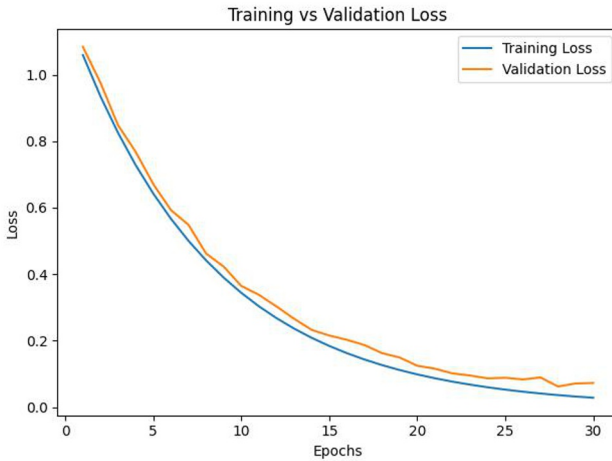


Fig. 4. Training and Validation Loss Curve

5. Conclusion & Future Scope

The proposed EdgeMambaViT-CNN architecture shows great results in the mango leaf disease detection with the proper combination of CNN-based local feature extraction, Vision Transformer-based global context learning, and Mamba-based efficient sequence modeling. The classification accuracy, interpretability and adaptability is further improved with integration of an attention-based feature fusion mechanism and an agentic decision layer. The findings of the experiment, such as high accuracy, good precision-recall performance, and good behavior during training, verify that the model is robust and fit to be applied in the real-world scenario (agriculture) especially when used on edge devices like smartphones.

To be applied in the future, the model can be expanded to allow real-time field deployment with the integration of the IoT, allowing automated disease monitoring in large farms. Moreover, further generalization can be achieved by using bigger and more diverse datasets, such as multi-crop and multi-disease scenarios. Self-supervised learning, domain adaptation, and federated learning could be used to improve performance in low-data settings due to the use of advanced methods. Moreover, explainability and addition of multimodal information (e.g. weather or soil conditions) might result in more intelligent and complete precision agriculture decision-support systems.

Acknowledgments. This is to show their heartfelt appreciation to all the people and organizations that helped in the successful completion of this research work. Special credits are given to the supporting organizations and academic environment which enabled this study. No particular grant was provided to this study by either the public, commercial, or not-for-profit funding agencies.

Disclosure of Interests. The authors (Vivek Kumar and Dr. Satender Kumar) state that they do not have any competing interests to disclose, which are pertinent to the content of this article.

References

1. Rao, U.S., et al.: Deep learning precision farming: Grapes and mango leaf disease detection. *Global Transitions Proceedings* 2(2), 535–544 (2021)
2. Manoharan, N., et al.: Identification of mango leaf disease using deep learning. In: Proc. ASIANCON, pp. 1–8 (2021)
3. Saleem, R., et al.: Mango leaf disease recognition using deep learning. *Applied Sciences* 11(24) (2021)
4. Tusher, A.N., et al.: Automatic recognition of plant diseases using CNN. In: ICIPCN, pp. 130–142 (2022)
5. Thaseentaj, S., Ilango, S.S.: Transfer learning for plant disease detection. *Computers, Materials & Continua* 77, (2023)
6. Hossain, M.A., et al.: Vision transformer for mango leaf disease identification. *Heliyon* (2024)
7. Rani, A., Gowrishankar, S.: ConvNeXt-based plant disease detection. *IJACSA* 14(6) (2023)
8. Puranik, S., et al.: MobileNetV3 for plant disease detection. In: Proc. INCET (2024)
9. Islam, M.M., et al.: Efficient sequence modeling for image classification. *Ecological Informatics* (2025)
10. Pratap, V.K., Kumar, N.S.: Deep learning-based mango leaf disease classification. *Fusion: Practice and Applications* 15 (2024)
11. Rao, U.S., et al.: Deep learning precision farming: Grapes and mango leaf disease detection by transfer learning. *Global Transitions Proceedings* 2(2), 535–544 (2021)
12. Manoharan, N., et al.: Identification of mango leaf disease using deep learning. In: Proc. ASIANCON, pp. 1–8 (2021)
13. Saleem, R., et al.: Mango leaf disease recognition and classification using segmentation and vein pattern techniques. *Applied Sciences* 11(24) (2021)
14. Tusher, A.N., et al.: Automatic recognition of plant leaf diseases using deep learning and image processing. In: ICIPCN, pp. 130–142 (2022)
15. Rani, A., Gowrishankar, S.: ConvNeXt-based mango leaf disease detection. *IJACSA* 14(6) (2023)
16. Rizvee, R.A., et al.: LeafNet: A CNN-based model for mango leaf disease detection. *Journal of Agricultural and Food Research* 14 (2023)
17. Thaseentaj, S., Ilango, S.S.: Deep CNNs for mango leaf disease detection. *Computers, Materials & Continua* 77 (2023)
18. Hossain, M.A., et al.: Deep learning for mango leaf disease identification using vision transformers. *Heliyon* (2024)
19. Puranik, S., et al.: MobileNetV3 for mango leaf disease detection. In: Proc. INCET (2024)
20. Recent study on YOLO-based mango leaf disease detection achieving high accuracy (YOLOv8/v9) (2024)
21. Ahmmed, J., et al.: Fine-tuned CNN-based approach for multi-class mango leaf disease detection. *arXiv* (2025)
22. Islam, M.M., et al.: Two-stage model for enhanced mango leaf disease detection. *Ecological Informatics* (2025)
23. Khandelwal, S., et al.: Optimizing performance in mango plant leaf disease detection. *ETASR* (2024)
24. Mishra, U., et al.: Deep learning-based disease detection in potato and mango leaves. *Scientific Reports* (2025)
25. Pratap, V.K., Kumar, N.S.: Deep learning-based mango leaf disease detection and classification. *Fusion: Practice and Applications* 15 (2024)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

