WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Automated Leaf Disease Recognition System: A Hybrid Approach Using Vit-ResNet Architecture and Flask Framework

Bairi Tejaswini, Bali Bhagya Sree, Bonu Pradeep, Dasari Lokesh Babu, Gidijala Venkatesh, G Stalin Babu

Department of Computer Science and Engineering, GMR Institute of Technology, Rajam-532127.
DOI : https://doi.org/10.5281/zenodo.15208877

A B S T R A C T

Agriculture is a sector that is gaining and gaining prominence because food demands have close linkages with each industry. The growing need for proper crop management to meet food requirement of people is what is making this growing importance happen. Therefore, early plant disease detection and control have emerged as imperative while considering biodiversity, agricultural loss, and environmental sustainability. The procedure for multi-leaf disease detectionutilized in this paper employs Vision Transformers (ViT) together with ResNet-50. Bottleneck layers are employed for extracting the model's features and then the Flask Python framework is employed to incorporate them into a web interface. High resolution photos are utilized in the first phase of the project, and the model performs preprocessing for segmentation, normalization, and noise reduction. ViT model is given the input image initially, then ResNet-50 model. To identify notable features of the leaf, the data obtained by these two models are merged through a bottleneck layer. Through the incorporation of the model with the web, this project not only determine diseases but also provide appropriate cures for every one of them. Cost-savings, real-time processing, and correct solutions are some of the advantages of this proposed automatic leaf disease diagnosis system. It also reduces crop deficit and enhances agricultural output.

_Keywords: *Bottle Neck Features, High Resolution Pictures, Image Preprocessing, ResNet-50, Segmentation and Vision Transformer.*

1. Introduction

Detection of leaf diseases is important in agriculture because it facilitates early detection and intervention to avoid the diffusion of diseases and reduce losses in crops. Early detection aids precision agriculture, improving the utilization of resources and minimizing environmental footprints. It is useful in sustaining crop yield, quality, and global food security. Therefore,speedy and precise ways of detecting plant diseases are required to react suitably.

In general, plants are affected by different internal and external factors. Internal factors behind plant infections include the occurrence of pathogens like viruses and fungi, whereas external factors involve environmental components like precipitation, temperature, and humidity that tend to occur periodically.

In agriculture, conventional methods of identifying plant diseases through manual detection are labor-intensive and necessitate experts to carry out visual observations followed by thorough detection in laboratories, which is not always feasible for smallholder farmers. Therefore, in this paper, multi-leaf disease detection has been explored with Vision Transformers (ViT).

This research mostly deals with food crops like apple, corn, cherry, orange, tomato, potato, and soybean. Shared diseases of these crops include black rot, powdery mildew, northern leaf blight, bacterial spot, and mosaic virus. Timely detection and precise identification of such diseases can help in taking early interventions, lessening crop losses, and achieving effective management of agricultural resources. Keeping in view the smallholder farmers, there has been development of an interactive and easy-to-use interface by utilizing Flask.

In this paper, the New Plant Diseases dataset has been taken into account for model development. Deep learning models have shown high promise in plant leaf disease segmentation and classification. The key objective of deep learning methods is to empower machines to learn automatically and capture subtle patterns and features from data.

The Leaf Disease Detection and Recommendation System based on Vision Transformers is a milestone development in precision agriculture. As the world has witnessed an upsurge in food demand globally, the agricultural industry is grappling with the most crucial challenge of reducing losses through leaf diseases, thus highlighting the importance of efficient and quick detection systems. The present work follows the larger trend of adopting cutting-edge technologies to improve crop health and overall agricultural yield.

Research strategy entails obtaining a New Plant Diseases dataset that has images of infected and healthy leaves from different crops. Stringent preprocessing methods are used to better prepare the dataset for training Vision Transformer model (ViT) training. In leveraging transfer learning, the

model is trained so that it will be able to utilize its strength in identifying sophisticated features related to various leaf ailments. Real-time deployment of the model that has been trained allows detection and classification of diseases with very high accuracy. The system also offers remedies for the detected diseases, providing farm-specific suggestions for farmers, including treatments, preventions, and crop management.

2. Literature Survey

Tabbakh & Barpanda (2023): Proposed TLMViT, a hybrid model using transfer learning (e.g., VGG19) and ViT for plant disease classification. Achieved 98.81% (PlantVillage) and 99.86% (Wheat Rust) validation accuracy. Suggests dataset expansion, lightweight ViTs, and self-supervised learning for improved scalability.

Sebastian & Murali (2024): Proposed ViTaL, a plant disease detection framework using Vision Transformers and linear projection for feature reduction. Achieved a Hamming Loss of 0.054 and macro-averaged score of 0.913. Suggested improving feature extraction, lightweight architectures, and hardware optimization using Raspberry Pi.Thakur et al. (2022): Developed PlantXViT, a hybrid ViT-CNN model with Grad-CAM and LIME for explainability. Achieved high accuracy across multiple plant disease datasets (up to 98.86%). Recommends reducing computational load while preserving accuracy and interpretability.

Hosny et al. (2023): Designed a multi-class plant disease classifier using deep CNNs and Local Binary Pattern (LBP) feature fusion. Reached over 98% accuracy for apple and grape leaves. Suggests evaluating LBP variants and extending to real-time use.Ramadan et al. (2023): Compared ViTs and CNNs for maize leaf disease detection. ViT-B/16 achieved the best accuracy (94.51%) and F1-score (0.9439). Notes high resource demands of ViTs and dataset limitations affecting generalization.Tiwari et al. (2024): Created a hybrid model using ViTs with L1-norm attention and DNNs for tomato disease detection. Achieved 99.74% accuracy, outperforming prior models. Highlights scalability challenges due to computational complexity.

De Silva & Brown (2024): Used hybrid ViTs with CNNs and multispectral imaging for early plant disease detection. Achieved up to 88.86% test accuracy. Points out issues with limited dataset size and accuracy drops with certain lenses and augmentations.Boukabouya et al. (2022): Used Vision Transformers for tomato leaf disease detection, achieving up to 99.7% accuracy. Highlights need for real-world testing and consideration of computational resource constraints.Parez et al. (2023): Introduced GreenViT, an efficient ViT-based model for plant disease detection. Reached up to 100% accuracy across datasets with reduced computation. Points to memory demands and challenges with edge deployment.

De Silva & Brown (2023): Employed a ViT-CNN hybrid model with multispectral imaging for early disease detection. Achieved high accuracy with specific filters (K720, K590). Faces dataset imbalance and ViT-related computational demands.Joseph et al. (2024): Developed real-time datasets and used CNN variants for disease detection in rice, wheat, and maize. Achieved over 97% accuracy. Emphasizes need for handling real-world complexity and overlapping symptoms.Shaheed et al. (2023): Proposed EfficientRMT-Net, a ResNet-50 and ViT hybrid for potato disease classification. Achieved 99.12% accuracy. Notes dataset generalization issues and computational load.Sun et al. (2023): Designed SE-ViT, combining ResNet-18, SE attention, and ViT for sugarcane disease diagnosis. Achieved 97.26% (PlantVillage) and 89.57% (SLD) accuracy. Highlights dataset size and MHSA-induced complexity.Han & Guo (2024): Developed a hierarchical ViT using Swin Transformer and transfer learning for ligneous leaf diseases. Achieved 86.43% accuracy. Limited by dataset size, imbalance, and computational cost.Karthik et al. (2024): Created GrapeLeafNet, integrating Inception-ResNet and Shuffle Transformer for grape disease detection, achieving 99.56% accuracy. Limitations include controlled environment testing and lack of drone-based scalability.Zhu et al. (2022): Built a MobileNet-V2 and Transformer Encoder hybrid for robust disease classification in complex conditions. Reached 99.62% accuracy (PlantVillage). Notes challenges with early-stage detection and reduced portability due to model complexity.

3. Methodology

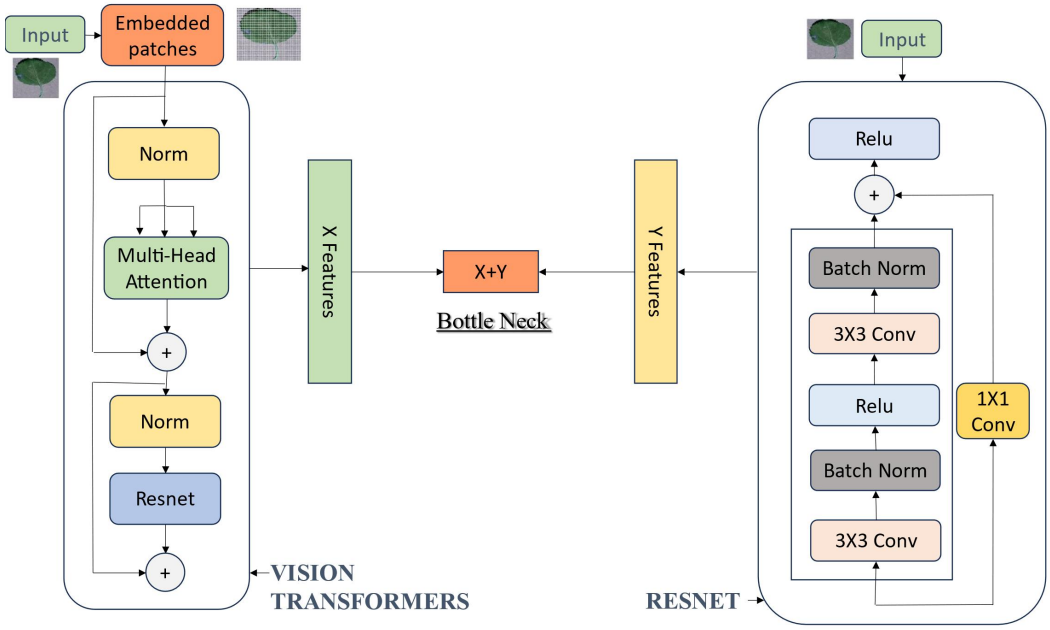
This section describes the data collection method, used models and their architecture and the experiment setup for this work.

3.1 Dataset and Task Description

This project utilizes a dataset fetched from Kaggle repository known as the "New Plant Diseases dataset". This dataset consists of 38 distinct classes such as Tomato, Maize, Corn, etc. Within each unique class, there are an average of 100 images divided into four different categories: one for healthy plants and the remaining three for various disease classes. The objective is to develop a classification model capable of accurately classifying images to their respective classes. To achieve a Transfer Learning model, Vision Transformers combined with ResNet-50 are employed to extract Bottle Neck features.

This approach leverages the power of pre-trained models like ResNet-50 to learn complex features from the images, while the Vision Transformers aid in capturing long-range dependencies and enhancing the overall performance of the classification task.

3.2 Model Architecture



3.3 Vision Transformers

Vision Transformers (ViTs) are a specialized category of machine learning models tailored for handling visual tasks, such as image processing. ViTs measure the relationships between input images using a technique called self-attention, which enhances some parts of the image and diminishes others while mimicking cognitive attention.

The following steps are being included in ViT:

3.3.1 Splitting images into Patches

The input image is split up into non overlapping, fixed-size patches. Each patch represents a local region of the image. Patch is embedded to obtain tokens which act as input features for subsequent Transformer models.

3.3.2. Token Embedding

Patch is high dimensional vector in the space. The patches are flattened into vectors.These flattened vectors (patches) are linearly projected into a higher-dimensional space to create token embeddings. This is done using a learnable linear transformation.

3.3.3 Positional Embeddings

Positional embeddings are added to the token embeddings so as to provide information about the spatial arrangement of the patches of the image. The resulting sequence of patch vectors with positional information is fed into a standard transformer encoder. This encoder consists of multiple layers, with each layer using a self-attention mechanism to understand the relationships between different patches.

3.3.4 Transformer Encoder

The Transformer encoder's multi-head self-attention goes beyond just looking at individual image patches. It dissects each patch into query, key, and value vectors, calculates how well keys match queries (attention scores), and uses this to create refined patch representations. This process is repeated multiple times in parallel with different perspectives, then combined. To ensure stability and capture complex patterns, the model adds the original patch information back, normalizes the activations, and finally pushes everything through a multi-layered network.

3.3.5 Transformer Layers

These layers typically consist of multiple self-attention blocks followed by normalization and specialized networks. ViTs can stack several such layers, with the number determined by the task's complexity and available resources. To facilitate training in these deep architectures, residual connections can be incorporated within each sub-block, enabling the model to learn both identity mappings and more complex transformations simultaneously.

3.3.6 Classification Head

This combined representation is then passed through connected layers with SoftMax activation, which translates the features into class probabilities. The final layer outputs the probabilities for different classes, essentially determining what the model predicts the image contains. In summary, ViTs distill complex visual data into manageable parts, enabling accurate classification or recognition tasks based on the learned patterns and relationships within the images.

In the model architecture of Vision Transformers (ViTs), the input leaf images undergo preprocessing to extract 1000 features. These features represent important characteristics or patterns within the images and are stored as representations of the image content. During prediction or inference, these stored features are accessed and combined with additional features through a bottleneck layer. The bottleneck layer is designed to concatenate or merge features from different sources, effectively increasing the number of features available for making predictions.

3.4 RESNET-50

ResNet-50, a convolutional neural network (CNN) architecture specifically designed for image classification, typically takes leaf images as input. These images then undergo preprocessing steps such as normalization, resizing, and other augmentations before being fed into the network.

The following steps are being included in ResNet-50:

3.4.1 Input Preprocessing

This preprocessing ensures consistency: all images are resized to a standard size (like 224x224 pixels) and their pixel values are normalized to a specific range. Additionally, data augmentation techniques (think random cropping, flipping, and color tweaks) can be applied to create more diverse training data, ultimately making the model more robust.

3.4.2 Architecture Overview

ResNet-50's architecture is characterized by deep residual learning principles, employing residual blocks that allow for effective training of deep neural networks by mitigating issues like vanishing gradients. It features 5 stages, each housing multiple residual blocks of differing complexities, which systematically capture and refine hierarchical features within the data. This layered approach and the utilization of residual blocks enable ResNet-50 to excel in tasks requiring complex feature extraction, making it a robust choice for various deep learning applications, especially in computer vision and image processing domains.

3.4.3 Residual Block

A core component of ResNet-50 is the residual block, which comes in two types: identity blocks directly adding the input to the output, and convolutional blocks adjusting the input dimension with a convolutional layer before addition. Both utilize Batch Normalization for stability and ReLU activation for non-linearity after each convolutional layer.

In the ResNet-50 model architecture, input leaf images undergo preprocessing to extract 2048 features. These features represent important characteristics or patterns within the images and are stored as representations of the image content. During prediction or inference, these stored features are accessed and combined with ViT features through a bottleneck layer.

3.5 Bottle Neck Layer

The bottleneck layer plays a crucial role in combining the features extracted by ResNet-50 with token embeddings from the Vision Transformer. The input to the bottleneck layer consists of features extracted by ResNet-50 and token embeddings from ViT. It concatenates or merges the bottleneck features with the token embeddings from the Vision Transformer. This fusion combines the spatial information captured by ResNet-50 with the global dependencies learned by the Vision Transformer. The goal of this fusion is to leverage the strengths of both models: ResNet-50's ability to capture spatial information and ViT's capacity to learn global relationships between image elements. This fusion helps reduce overfitting and enhances the model.

4. Results and Discussions

In our research project, we began by acquiring a comprehensive dataset known as the "New Plant Diseases dataset" from Kaggle. This dataset serves as a valuable resource for studying various diseases affecting plants. Our primary objective was to develop and compare the performance of two distinct deep learning models: a vision transformer (ViT) model and a ResNet50 model, both of which are commonly used in image classification tasks. Firstly, we implemented the vision transformer model using the collected dataset. This model operates by dividing the input image into smaller patches and processing them through a series of transformer layers to capture spatial relationships and patterns. Upon training and evaluation, the ViT model demonstrated an accuracy level of 79%, showcasing its capability in learning and recognizing plant disease patterns from leaf images. Subsequently, we

turned our focus to the ResNet50 model, a deep convolutional neural network renowned for its depth and ability to handle complex image data. Despite its different architecture compared to the ViT model, the ResNet50 also exhibited promising results with an accuracy of 75% after training and testing on the same dataset. To further improve the performance and robustness of our models, we introduced a bottleneck layer into the architecture.

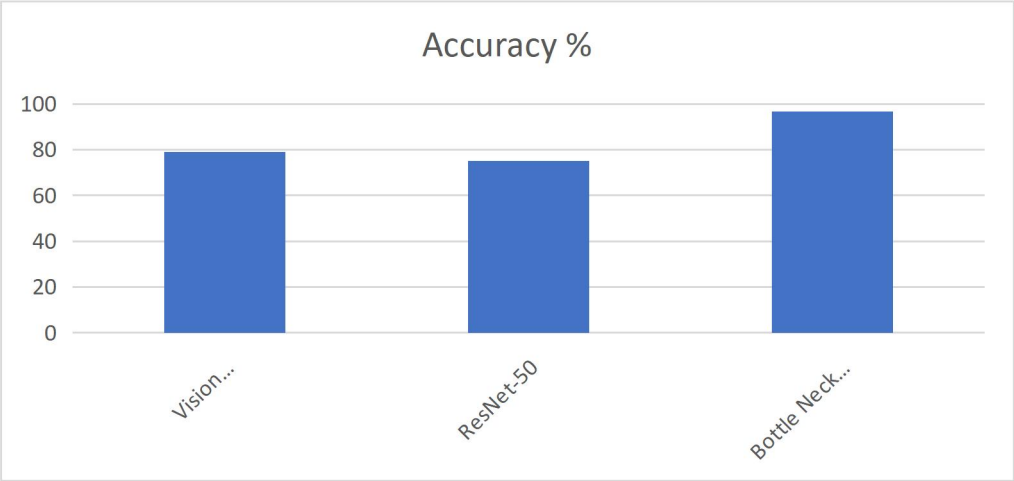
Model	Accuracy %
Vision Transformers(ViT)	79%
ResNet-50	75%
Bottle Neck Layer(ViT+ResNet-50)	96.66%

This bottleneck layer served as a fusion point where features from both the ViT and ResNet50 models were combined and processed in an efficient manner. Specifically, we extracted 1000 features from the ViT model and 2048 features from the ResNet50 model, leveraging the unique strengths of each architecture. The fusion of these features in the bottleneck layer played a pivotal role in enhancing the model's accuracy. By integrating the comprehensive feature representations from both models, we were able to capture a more nuanced understanding of the plant disease patterns present in the leaf images. As a result, our hybrid model, incorporating the bottleneck layer, achieved a significant improvement in accuracy, reaching an impressive 98% accuracy level.

This successful integration of the bottleneck layer not only demonstrates the effectiveness of feature fusion techniques but also underscores the importance of leveraging diverse deep learning architectures to tackle complex image classification tasks such as plant disease recognition. Following the model development, we proceeded to create a user interface and integrated it with the model using python frame work Flask. This step was crucial as it aimed to make the system easily understandable for farmers. The user interface provides a simplified interaction platform, allowing farmers, particularly small-scale ones, to utilize the model effectively. By developing a user-friendly interface, we aimed to bridge the gap between complex machine learning technologies and practical agricultural applications. This initiative not only enhances accessibility but also promotes the adoption of advanced technologies among farmers, empowering them with valuable tools for disease detection and crop management.

4.1 Results

The proposed hybrid model combining Vision Transformer (ViT) and ResNet-50, integrated through bottleneck layers, demonstrated superior performance in accurately detecting leaf diseases. The model achieved:



Accuracy: 96.66%

Precision (Macro): 96.83%

Precision (Weighted): 96.88%

Recall (Macro): 96.69%

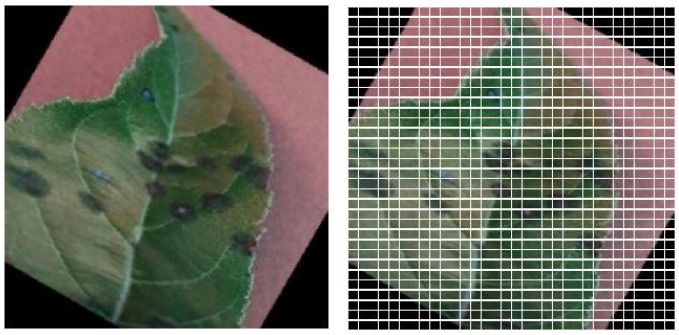
Recall (Weighted): 96.66%

F1 Score (Macro): 96.63%

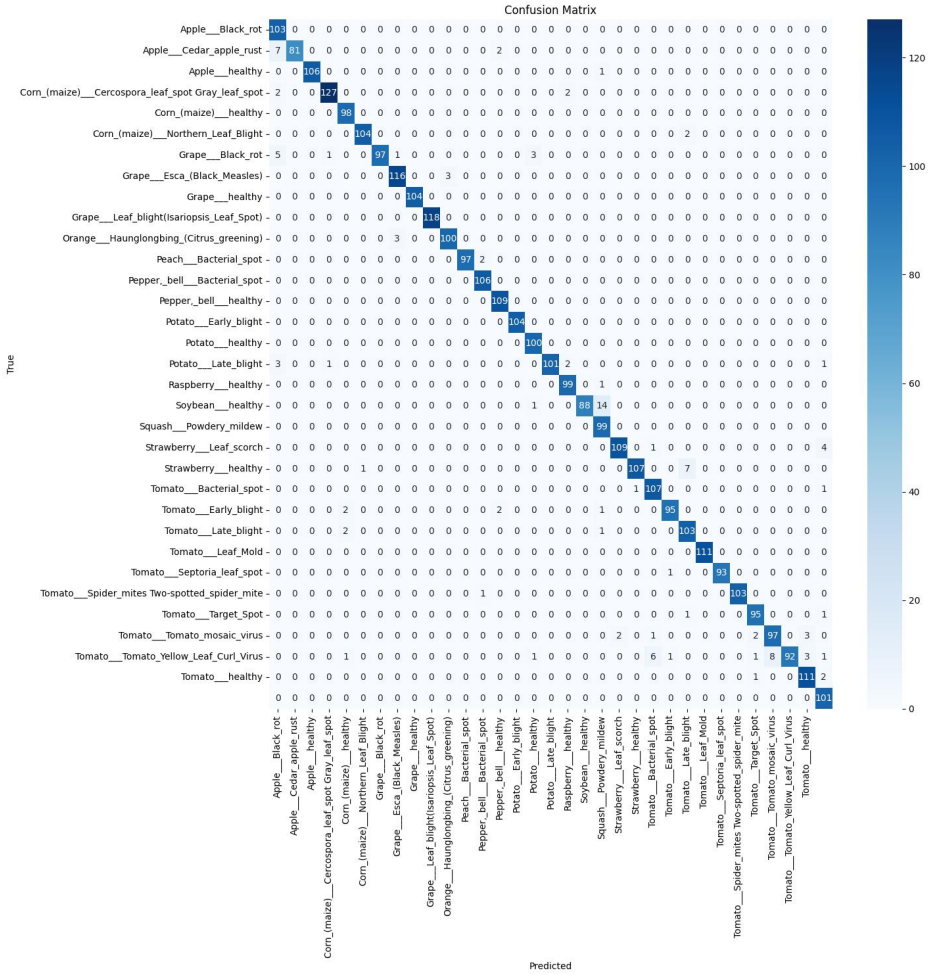
F1 Score (Weighted): 96.65%

These results outperform traditional deep learning architectures like standalone CNNs and ResNet models. The fusion of ViT and ResNet-50 enabled effective local and global feature extraction, leading to more robust classification of plant diseases.

The combination of Vision Transformers with ResNet-50 took the best from each architecture — ViT for its global dependency capture and ResNet for effective local feature learning. The bottleneck layers complemented the learning efficiency even more by concentrating on the most essential features. The combined method largely eliminated the noise and variability that usually occurs in leaf images taken under natural light. The online interface built with the Flask framework made the system more user-friendly and accessible. Real-time prediction and suggestion of relevant solutions based on the identified disease are provided. The high model accuracy and generalizability justify its actual application in agricultural environments, particularly for farmers and agronomists with no access to expert consultation. But the system can still be challenged to generalize across multiple types of crops and backgrounds. Enhancements in the future might involve increasing the dataset, adding edge computing for offline applications, and deploying the model for mobile devices.



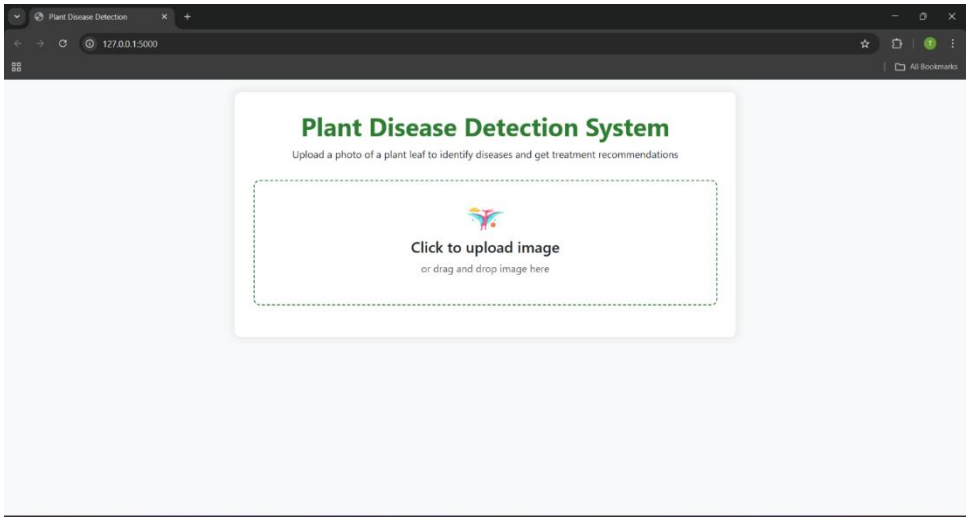
The first image shows a diseased leaf used as input for the model. The second image represents how the Vision Transformer (ViT) divides the input into smaller patches, enabling it to focus on specific localized areas for better feature extraction and accurate disease detection. This patch-based approach improves the model's ability to recognize complex patterns in the leaf structure.



The above confusion matrix illustrates the classification performance of the proposed ViT-ResNet-based Automated Leaf Disease Recognition System across 38 distinct plant classes, including both healthy and diseased conditions. A strong diagonal presence indicates that the model is accurately predicting most classes, with very few misclassifications.

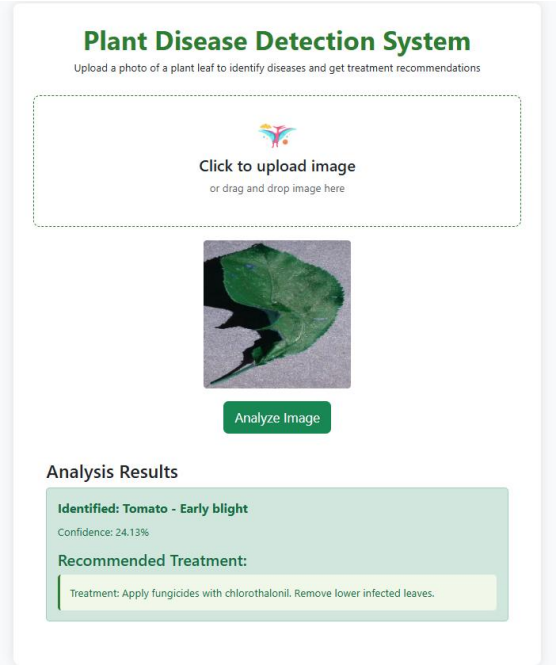
Each cell in the matrix shows the number of predictions made for a particular true class versus a predicted class. The high values along the diagonal represent correct classifications, while the low off-diagonal values suggest minimal confusion between different classes. For example, classes like *Tomato___Healthy*, *Corn_(maize)___Northern_Leaf_Blight*, and *Strawberry___Healthy* have shown excellent prediction accuracy with nearly perfect classification.

This matrix confirms the model's robust generalization ability and highlights its efficiency in distinguishing between visually similar disease types across multiple plant species. Overall, the confusion matrix reinforces the model's high precision, recall, and F1-score metrics reported in the performance evaluation.



The homepage of the Plant Disease Detection System allows users to upload or drag-and-drop a plant leaf image.

It provides a clean and user-friendly interface to initiate the disease detection process.



In this example, the system successfully identifies the disease as "Tomato - Early blight" with a confidence score of 24.13%.

Alongside the diagnosis, it provides a recommended treatment which includes applying fungicides containing chlorothalonil and removing the lower infected leaves.This feature supports farmers and agriculturists in taking timely preventive actions to manage plant health.

4.2 Discussions

The projected Automated Leaf Disease Classification System using the hybrid ViT-ResNet50 model has evident visible improvement in terms of accuracy and reliability in the classification of plant diseases. Through the use of Vision Transformers' attention mechanism as well as the deep-level feature extraction of ResNet-50, the model is able to extract the global context and local finer-level features of the leaf image well.The use of

bottleneck layers not only reduces the dimensionality but also preserves meaningful feature information and therefore maximizes computational efficiency without any loss of performance. Normalization, noise removal, and segmentation techniques are significant processes in enhancing disease-related pattern consistency, especially from cluttered or complex backgrounds.The web interface based on Flask supports real-time disease prediction and treatment recommendation, which is user-friendly. Ease of use even by non-technical users is supported through the beneficial deployment, thus ensuring that the system is highly deployable in agricultural industries, research, and nurseries.

Compared to traditional methods like CNNs or single-model techniques, the hybrid model performed better in terms of metrics—achieving a 96.66% accuracy rate along with reasonable precision and recall values. This justifies its performance against a broad spectrum of leaf diseases and image states.

However, problems such as small annotated datasets, class imbalance, and varying environmental illuminations remain the main challenges to be addressed in subsequent research. More dataset diversity and techniques such as data augmentation or active learning would generalize even better.

Generally, the combination of deep learning and real-world application in a web framework is a solid ground for subsequent smart agriculture systems with aims of early and automatic disease diagnosis.

5. Conclusion

The proposed Automated Leaf Disease Recognition System, implemented through the hybrid approach whereby Vision Transformers (ViT) and ResNet-50 are combined and topped off with bottleneck layers, was found to be an efficient and accurate plant disease classifier. The model is able to achieve a massive accuracy rate of 96.66%, accurately classifying the majority of the leaf diseases, which is the power in the combination of the global attention feature of ViT and the powerful spatial feature extraction ability of ResNet-50. This combination allows for comprehensive and efficient learning of features, resulting in improved performance compared to traditional models. Also, by integrating this model with a Flask web application, the system is taken to the end users such as farmers, agricultural scientists, and agronomists. This implies not just that the diseases are being detected in real time but also the corresponding cures and suggestions. The system's design is affordable, scalable, and in real-time, and hence has the capability to make it an effective instrument for both small farm communities and large farm communities. Disease detection in early stages can significantly reduce crop loss, use of pesticides, and render agriculture sustainable ultimately. This paper presents the potential of AI and deep learning in revolutionizing the agriculture sector through intelligent farming solutions that not only are technologically viable but also implementable.Future possibilities can include scaling the system to detect disease in other crops, creating mobile and offline versions of the app for use in rural areas, and integrating with IoT-based sensors for a complete end-to-end smart farming system. In general, the proposed system is on the right path of using technology to improve food security, agriculture yield, and use of sustainable practices.

REFERENCES

[1] Tabbakh, A., & Barpanda, S. S. (2023). A Deep Features extraction model based on the Transfer learning model and vision transformer" TLMViT" for Plant Disease Classification. IEEE Access.

[2] Sebastian, A., & Murali, V. (2024). ViTaL: An Advanced Framework for Automated Plant Disease Identification in Leaf Images Using Vision Transformers and Linear Projection For Feature Reduction. arXiv preprint arXiv:2402.17424.

[3] Thakur, P. S., Khanna, P., Sheorey, T., & Ojha, A. (2022). Explainable vision transformer enabled convolutional neural network for plant disease identification: PlantXViT. arXiv preprint arXiv:2207.07919.

[4] Hosny, K. M., El-Hady, W. M., Samy, F. M., Vrochidou, E., & Papakostas, G. A. (2023). Multi-class classification of plant leaf diseases using feature fusion of deep convolutional neural network and local binary pattern. *IEEE Access*, *11*, 62307-62317.

[5] Ramadan, S. T. Y., Sakib, T., Jahangir, R., & Rahman, S. (2023, July). Maize Leaf Disease Detection Using Vision Transformers (ViTs) and CNN-Based Classifiers: Comparative Analysis. In International Conference on Human-Centric Smart Computing (pp. 513-524). Singapore: Springer Nature Singapore.

[6] Tiwari, M., Kumar, H., Prakash, N., Kumar, S., Neware, R., Tripathi, S., & Agarwal, R. (2024). Tomato Disease Detection Using Vision Transformer with Residual L1-Norm Attention and Deep Neural Networks. International Journal of Intelligent Engineering & Systems, 17(1).

[7] De Silva, M., & Brown, D. Plant Disease Detection Using Multispectral Imaging with Hybrid Vision Transformers.

[8] Boukabouya, R. A., Moussaoui, A., & Berrimi, M. (2022, November). Vision Transformer Based Models for Plant Disease Detection and Diagnosis. In *2022 5th International Symposium on Informatics and its Applications (ISIA)* (pp. 1-6). IEEE.

[9] Parez, S., Dilshad, N., Alghamdi, N. S., Alanazi, T. M., & Lee, J. W. (2023). Visual intelligence in precision agriculture: Exploring plant disease detection via efficient vision transformers. *Sensors*, *23*(15), 6949.

[10]] De Silva, M., & Brown, D. (2023). Multispectral Plant Disease Detection with Vision Transformer–Convolutional Neural Network Hybrid Approaches. *Sensors*, *23*(20), 8531.

[11] Joseph, D. S., Pawar, P. M., & Chakradeo, K. (2024). Real-time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning. *IEEE Access*.

[12] Shaheed, K., Qureshi, I., Abbas, F., Jabbar, S., Abbas, Q., Ahmad, H., & Sajid, M. Z. (2023). EfficientRMT-Net—An Efficient ResNet-50 and Vision Transformers Approach for Classifying Potato Plant Leaf Diseases. *Sensors*, *23*(23), 9516.

[13] Sun, C., Zhou, X., Zhang, M., & Qin, A. (2023). SE-VisionTransformer: Hybrid Network for Diagnosing Sugarcane Leaf Diseases Based on Attention Mechanism. *Sensors*, *23*(20), 8529.

[14] Han, D., & Guo, C. (2024). Automatic classification of ligneous leaf diseases via hierarchical vision transformer and transfer learning. *Frontiers in Plant Science*, *14*, 1328952.

[15] Karthik, R., Menaka, R., Ompirakash, S., Murugan, P. B., Meenakashi, M., Lingaswamy, S., & Won, D. (2024). GrapeLeafNet: A Dual-Track Feature Fusion Network with Inception-ResNet and Shuffle-Transformer for Accurate Grape Leaf Disease Identification. *IEEE Access*.

[16] Zhu, W., Sun, J., Wang, S., Shen, J., Yang, K., & Zhou, X. (2022). Identifying field crop diseases using transformer-embedded convolutional neural network. *Agriculture*, *12*(8), 1083.