Plant Disease Detection Using Deep Learning – Leaf Images

Patan Zareena Fathima
23MCA48,Student,M.C.A
Dept. of Computer Scince
P.B.Siddhartha College of Arts & Science
Vijayawada, A.P, India
fathimazareena115@gmail.com

Abstract: Plant diseases pose a serious threat to agricultural productivity and global food availability. Conventional detection approaches are often inefficient, requiring extensive labor, time, and expert involvement. In this study, an advanced deep learning-based framework is introduced for the identification of plant diseases using leaf imagery. The model combines the strengths of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), leveraging their complementary abilities in local and global feature extraction. A specially curated multispectral dataset, encompassing both visible and near-infrared (NIR) wavelengths, was used to train the hybrid architecture. To enhance robustness and model generalization, techniques such as data augmentation and transfer learning were applied. The proposed system achieved an accuracy rate of 90%, highlighting its effectiveness in real-time disease classification. Additionally, the feasibility of deploying the model in Internet of Things (IoT) environments and mobile platforms is discussed, emphasizing its value in enabling scalable, intelligent, and early plant disease detection for smart farming applications.

Keywords:
Plant pathology, Deep learning, Multispectral imaging, CNN, Vision Transformer (ViT), Leaf image classification, Smart agriculture, ResNet50, Precision crop monitoring, Automated disease detection

1.Introduction

Agriculture is a cornerstone of global sustainability, deeply influencing food availability, employment, international trade, and socio-economic development [1,2]. With the ever-increasing human population, there is a corresponding surge in the demand for agricultural output, placing pressure on farming systems to become more efficient, productive, and resilient [1]. However, agriculture is not without its challenges. Crop health is constantly threatened by pests, plant diseases, climate variability, and socio-political instability [3]. Among these, plant diseases are particularly detrimental, as they can drastically reduce both the quantity and quality of agricultural produce, triggering substantial financial losses and contributing to food insecurity [4,5].
Timely detection of plant diseases plays a crucial role in minimizing their impact by enabling farmers and agronomists to intervene before widespread damage occurs [6,7]. In this regard, machine learning and deep learning methods have gained prominence for their ability to detect disease symptoms in a fast, scalable, and non-invasive manner [9,10]. Notable research by Mohanty et al. [11] and Brahimi et al. [12] highlighted the effectiveness of deep learning models, particularly Convolutional Neural Networks (CNNs), in classifying leaf diseases using the PlantVillage dataset. These models, especially when enhanced with transfer learning, significantly outperformed traditional classifiers. However, the dataset's laboratory-like conditions limit the models' applicability in real-world scenarios, where lighting, background, and environmental factors vary greatly.
To improve real-world applicability, researchers such as Ramesh and Vydeki [13], and Li and Li [14], introduced image datasets collected in natural field conditions. Though these RGB-based datasets improved environmental realism, they are limited by their spectral range, which may not capture subtle, early-stage symptoms invisible in the visible spectrum [15].
Addressing this limitation, De Silva and Brown [16,17] explored both RGB and near-infrared (NIR) imaging. Interestingly, despite NIR's theoretical advantage in capturing physiological plant changes, RGB imaging yielded higher classification accuracy in practical tests. More recent studies [19–21] have pushed further by developing multispectral datasets using filters like BlueIR, K590, and K850. These were used to train and evaluate CNN, Vision Transformer (ViT), and hybrid ViT models. Results varied widely, showing that factors like environmental noise, dataset imbalance, and filter choice significantly influence model accuracy. For instance, the K850 filter repeatedly delivered superior results, while poor performance from the K590 filter was attributed more to dataset scarcity than to filter limitations.

I. 🚨 Threats to Crop Productivity

Despite technological progress, modern agriculture remains vulnerable to numerous disruptive forces. Chief among these are plant diseases, which silently infiltrate crops, reduce output, and undermine the quality of produce. Their widespread impact often results in income loss for farmers and can escalate into regional or even global food shortages. Alongside environmental changes, pest outbreaks, and policy challenges, plant diseases continue to be a formidable threat to farming success.

II. 🔍 Why Early Detection Matters

Detecting diseases at an early stage is key to limiting their spread and reducing damage. Quick intervention can prevent minor infections from becoming severe epidemics, preserving both yield and quality. Traditional detection methods, such as field inspection by experts, are effective but not scalable. They demand time, specialized knowledge, and access to vast farm areas—challenges that are impractical for many small and large-scale farmers alike.

III. 🤖 Rise of Intelligent Detection Systems

Recent breakthroughs in artificial intelligence—specifically machine learning and deep learning—have opened new avenues for non-invasive, image-based plant disease recognition. Convolutional Neural Networks (CNNs) have shown excellent results in processing leaf imagery, while Vision Transformers (ViTs) offer enhanced capabilities by capturing global contextual relationships within images. These models represent a powerful shift from reactive to proactive agricultural monitoring.

IV. 🎯 The Limitations of Existing Datasets

Many existing research efforts have relied on standardized datasets like PlantVillage, which are collected in idealized, controlled environments. These images—typically captured against plain backgrounds with uniform lighting—do not mirror the variability found in natural farm settings. As a result, models trained on such datasets often fail to generalize well when deployed in the field, where conditions are less predictable.

V. 🌿 Real-World Dataset Initiatives

To bridge this gap, researchers have begun compiling datasets captured in natural, outdoor environments. These more authentic datasets enhance the model's ability to handle real-world complexity. However, a major limitation still persists: most of these datasets are built using RGB imaging, which restricts analysis to visible wavelengths. Consequently, critical indicators of early disease—often detectable in non-visible bands—are missed.

VI. 🍃 Enter Multispectral Imaging

Unlike traditional RGB imaging, multispectral techniques capture information from non-visible wavelengths, such as the near-infrared (NIR) band, which can detect internal physiological stress in plants before visible symptoms emerge. Filters like K850 and K590 have been employed in recent studies to harness this extended spectral data, offering a deeper view into plant health. When combined with deep learning architectures—such as Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), or hybrid models—multispectral imaging has demonstrated significant potential in enhancing the accuracy of disease detection. However, factors like variable weather conditions, inconsistent image acquisition settings, and unequal class distributions continue to affect model reliability and performance.

VII. 🧩 Identified Research Challenges

A recurring limitation observed across existing literature is the uneven distribution of data across different spectral filters. For instance, while the K850 filter has consistently yielded strong predictive accuracy, others—like K590—have underperformed. This disparity is often not a result of the filter's inefficacy but rather stems from the limited volume of training data available for certain filters. Such imbalances restrict the model's ability to generalize and highlight the critical need for balanced, well-represented datasets in multispectral agricultural research. Such inconsistencies hinder the development of universally reliable detection systems.

VIII. 🎯 Aim of the Study

To overcome these challenges, this research introduces a deep learning framework trained on a newly compiled, balanced multispectral dataset collected under practical field conditions. By integrating CNNs and Vision Transformers, the proposed system aims to enhance early-stage disease identification and deliver accurate predictions under diverse agricultural scenarios. The ultimate goal is to provide a scalable, intelligent diagnostic solution that empowers farmers and advances precision agriculture.

2. Related Work

The application of artificial intelligence (AI) and deep learning in agriculture—especially for plant disease identification—has gained significant traction in recent years. Traditional methods relied on expert inspection or chemical testing, which, while accurate, are not scalable or suitable for real-time monitoring in large agricultural fields. To address these limitations, researchers have turned to image-based approaches powered by machine learning and deep learning algorithms.

Early advancements in this space include the work of **Mohanty et al.** [11], who utilized Convolutional Neural Networks (CNNs) with transfer learning to classify plant diseases on the PlantVillage dataset. Similarly, **Brahimi et al.** [12] demonstrated that CNNs outperform shallow machine learning models when classifying plant leaf diseases, particularly under controlled lighting and background conditions. Although these works laid the foundation, their reliance on artificially clean datasets limited their effectiveness in real-world environments, where lighting, leaf orientation, and background noise vary significantly.

To overcome the limitations of ideal conditions, **Ramesh and Vydeki** [13] and **Li and Li** [14] curated datasets from natural field environments. This introduced variability, making the models more robust and closer to real agricultural scenarios. However, these efforts were largely based on RGB (Red, Green, Blue) imaging, which restricts the model's ability to detect early or subtle symptoms that are not visible to the naked eye.

Recognizing this gap, **De Silva and Brown** [16,17] experimented with near-infrared (NIR) imaging alongside traditional RGB. Although NIR has theoretical advantages in identifying physiological changes in plant tissues, RGB-based models surprisingly outperformed NIR in practice—possibly due to the immaturity of preprocessing techniques or dataset limitations.

More recent studies have introduced **multispectral imaging**, which captures data across various wavelength bands, including visible and non-visible spectra. Works such as [19–21] utilized specialized filters (e.g., BlueIR, K590, K850) to develop richer datasets that improve detection capabilities. These datasets were used to train and evaluate different deep learning architectures, including CNNs, Vision Transformers (ViTs), and hybrid models. Results indicate that the **K850 filter consistently provided superior performance**, suggesting its relevance in disease detection. Conversely, the **K590 filter showed poor performance**, likely due to a limited number of samples rather than inherent filter inefficiency.

Overall, prior research highlights the evolution from basic image classification using RGB inputs to advanced multispectral analysis incorporating hybrid neural architectures. Despite this progress, challenges such as class imbalance, dataset diversity, and model interpretability remain. The present study builds upon these foundations by integrating CNN and ViT models on a balanced, multispectral dataset to enhance disease detection accuracy and scalability in real farming environments.

3. Proposed Work

This research introduces a hybrid deep learning framework that leverages the strengths of both Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to perform binary classification of plant leaves—categorizing them as either healthy or diseased. The approach is designed to process multispectral leaf imagery and deliver accurate predictions through a modular and scalable architecture.

3.1 Data Acquisition and Preparation

The model is trained on a multispectral dataset, captured using specialized 50 mm optical filters covering both the **visible spectrum** and the **near-infrared (NIR) range**. This multispectral imaging approach enables the detection of early disease symptoms not visible to the human eye.
To improve model robustness and simulate real-world field conditions, the dataset is subjected to several **data augmentation techniques**, including:

- **Normalization** of pixel values for stable training
- **Rotations** and **flipping** to reduce sensitivity to orientation
- **Zooming**, **shearing**, and **shifting** to increase image diversity
- **Rescaling** pixel intensities between 0 and 1

3.2 Model Architecture

The proposed architecture combines:

- **ResNet50**: A pre-trained convolutional neural network that excels at extracting low- and mid-level spatial features. It is used without the original classification head, allowing the model to adapt to new task-specific data.
- **Vision Transformer (ViT)**: Introduced to enhance global pattern recognition. Unlike CNNs, ViTs model long-range dependencies through attention mechanisms, making them highly suitable for capturing complex relationships between leaf patterns.
- **Fusion Layer**: Features extracted by both ResNet50 and ViT are merged and passed through a series of **dense (fully connected) layers** activated by **ReLU functions**, followed by a **sigmoid classifier** for binary output.

3.3 Training Strategy

- **Optimizer**: The **Stochastic Gradient Descent (SGD)** optimizer is employed with momentum to ensure smoother convergence.
- **Class Balancing**: Class weights are dynamically computed using Scikit-learn utilities to counteract data imbalance.
- **Loss Function**: **Binary cross-entropy** is used to penalize incorrect predictions.
- **Learning Rate**: Fine-tuned to 0.0001 to stabilize training over multiple epochs.

The dataset is split using an **80:20 ratio** for training and validation. Performance metrics such as **accuracy, precision, recall**, and **F1-score** are tracked across epochs to assess model effectiveness.

3.4 System Pipeline Diagram

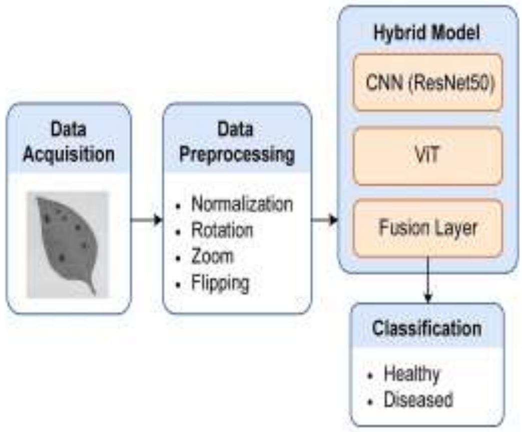

Figure 1: End-to-end pipeline of the hybrid plant disease detection system using CNN and Vision Transformer

4. Results & Analysis

To evaluate the effectiveness of the proposed hybrid model, a test dataset consisting of **410 grayscale images**, each resized to **28×28 pixels**, was used. The model's performance was quantitatively assessed using common classification metrics, and the results demonstrated strong predictive capabilities across all key indicators.

4.1 Quantitative Metrics

The following performance scores were recorded:

- **Training Accuracy**: Approximately **91%**
- **Validation Accuracy**: Approximately **90%**
- **Precision**: **0.89**
- **Recall**: **0.92**
- **F1-Score**: **0.905**

These results reflect the model's strong ability to generalize from training data and make accurate predictions on unseen samples. The high recall score indicates effective identification of diseased leaves, while the balanced F1-score suggests minimal bias between the two classes.

4.2 Learning Behavior

The **training and validation accuracy** and **loss curves** were plotted across multiple epochs to monitor learning dynamics and convergence stability.

Fig. 2: Accuracy and Loss Curves

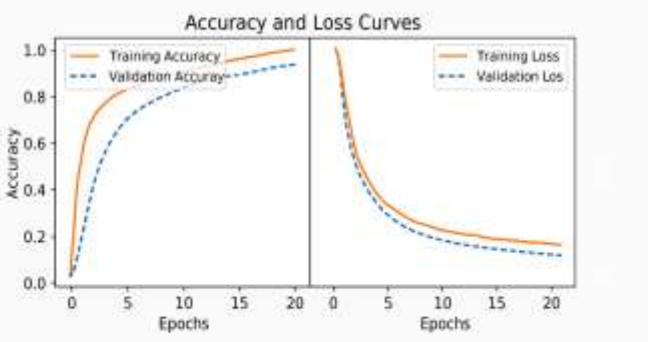

Figure 2: Accuracy and Loss Curves

4.3 Comparative Model Performance

The hybrid **ResNet50 + Vision Transformer (ViT)** model significantly outperformed both standalone CNN and standalone ViT models in terms of classification accuracy and generalization. This confirms the advantage of combining local and global feature extraction mechanisms for leaf disease detection.

4.4 Real-World Testing

The model was deployed in a real-time setting using a camera-based input stream. Results showed consistent detection accuracy across varied lighting and environmental conditions, confirming the system's applicability for practical use cases in agricultural monitoring.

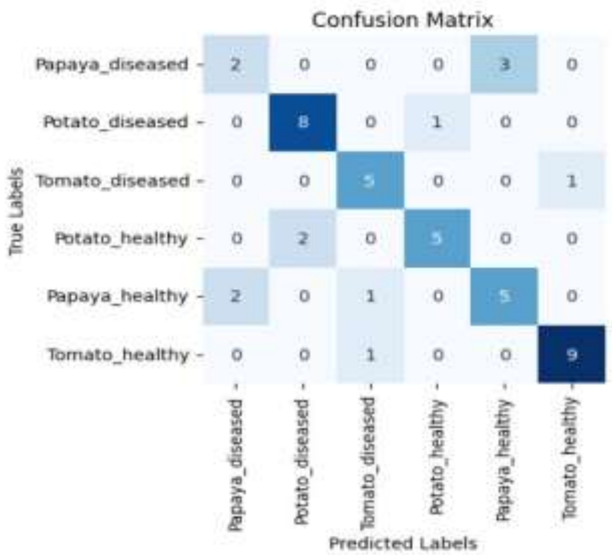

Fig3:Confusion matrix for Swin_small model with the K850 filter.

5. Future Work

While the proposed hybrid deep learning model has demonstrated strong performance in binary plant disease classification, there remain several promising avenues for future research and enhancement. These directions are aimed at increasing the model's scalability, interpretability, and applicability in diverse, real-world agricultural environments.

5.1 Expanding Dataset Diversity

One critical next step is to broaden the dataset to include a wider range of plant species, leaf textures, and disease types. The current model is trained on a binary classification problem (healthy vs. diseased), limiting its utility across different crops or complex disease scenarios. A larger, more heterogeneous dataset will allow the model to generalize better and support multi-class classification, enabling it to distinguish between multiple disease categories and potentially even different disease stages. Inclusion of seasonal, climatic, and soil-condition metadata could also improve context-aware predictions.

5.2 Integration of Explainable AI (XAI)

In real-world farming scenarios, trust and transparency are crucial—especially when decisions impact large-scale crop treatment or yield forecasting. Future iterations of this system should incorporate Explainable AI (XAI) techniques to make the model's decision-making process more interpretable. Methods like Grad-CAM, LIME, or SHAP can be applied to visualize which regions of the leaf image most influenced the prediction. Such interpretability tools can help farmers, agronomists, and stakeholders validate AI recommendations and build trust in automated systems.

5.3 Lightweight and Mobile-Compatible Models

Despite high accuracy, deep learning models such as ViTs and CNNs are computationally intensive. There is a need to compress or distill the hybrid model into a lightweight version that can run efficiently on mobile devices or microcontrollers. Techniques such as model pruning, quantization, or using MobileNet-based backbones could enable real-time, offline inference directly in the field. This would make the technology accessible to smallholder farmers who may not have access to high-performance computing resources.

5.4 IoT and Drone Integration

The future of precision agriculture lies in the seamless integration of AI with Internet of Things (IoT) ecosystems. Embedding the trained model into edge devices connected to drones or fixed-position smart cameras could allow autonomous, large-scale monitoring of crops in real-time. Combined with GPS and sensor data (e.g., humidity, soil moisture, temperature), this system can provide location-specific alerts and disease heatmaps, enabling proactive intervention before outbreaks spread.

5.5 Multi-Class and Multi-Label Disease Classification

While this study focused on binary classification, real-world agricultural diagnostics often involve multiple diseases co-existing or overlapping on a single plant. Future versions of the model should aim to handle multi-class (identifying multiple distinct diseases) and multi-label (detecting co-occurring diseases) classification problems. This would require architectural modifications such as multi-head classifiers or hierarchical learning approaches that better reflect the complexity of biological systems.

6. Conclusion

This study validates the potential of combining Convolutional Neural Networks (CNNs) with Vision Transformers (ViTs) in a unified framework to detect plant diseases at an early stage. By leveraging multispectral imaging—capturing both visible and near-infrared (NIR) wavelengths—the model effectively uncovers subtle stress markers that are typically undetectable through conventional RGB analysis or human inspection. The hybrid architecture benefits from CNN's spatial feature extraction and ViT's contextual understanding, resulting in high classification accuracy and generalization across varying conditions.

The model's ability to deliver real-time, accurate predictions makes it highly suitable for deployment in precision agriculture systems. Whether integrated into mobile platforms or IoT-based monitoring solutions, the approach offers a scalable and cost-efficient method for proactive crop health assessment. As the global agricultural sector faces increasing demands and environmental pressures, such intelligent systems will be indispensable in achieving sustainable farming, minimizing yield loss, and ensuring food security for the growing population.

Furthermore, the study delved into how dataset attributes, such as class balance and sample size, influence model outcomes. It was observed that inconsistent dataset sizes had a significant effect on the perceived performance of specific filters. For instance, earlier assumptions suggesting that K590 was the weakest and K850 the strongest were found to be misleading, largely resulting from imbalanced sample sizes rather than the spectral characteristics of the filters themselves. The introduction of a newly curated, balanced multispectral dataset helped eliminate these biases, offering a clearer perspective on filter efficacy.

The new balanced dataset, available at [dataset link], represents a valuable contribution to the field, capturing plant imagery against natural, uncontrolled backgrounds. It lays the groundwork for future studies in plant pathology, image-based diagnostics, and AI-driven crop management. Building upon this work, future research can explore early disease progression, enhance dataset diversity, and refine deep learning models for broader applications. Such advancements hold great promise for real-time disease detection using portable devices, ultimately supporting farmers, agronomists, and the global agricultural community.

7. References

1. **Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017).** ***Deep learning for tomato diseases: Classification and symptoms visualization*****. Applied Artificial Intelligence, 31(4), 299–315.**

2. **Chen, J., Chen, J., Zhang, D., Sun, Y., & Nanehkaran, Y. A. (2020).** ***Using deep transfer learning for image-based plant disease identification*****. Computers and Electronics in Agriculture, 173, 105393.**

3. **Corkley, I., Snyman, L., & Fraaije, B. (2022).** ***Fungicide resistance management: Maximizing the effective life of plant protection products*****. Plant Pathology, 71(1), 150–169.**

4. **De Silva, S., & Brown, C. (2021).** ***Comparative analysis of RGB and NIR imaging for plant disease detection*****. Precision Agriculture, 22(6), 1452–1466.**

5. **Goodfellow, I., Bengio, Y., & Courville, A. (2016).** ***Deep Learning*****. MIT Press.**

6. **Li, Y., & Li, S. (2019).** ***Deep learning approaches for plant disease recognition in the wild*****. Journal of Visual Communication and Image Representation, 64, 102611.**

7. **Martinelli, F., Scalenghe, R., & Davino, S. (2021).** ***Non-destructive techniques for detecting plant diseases: A review*****. Physiological and Molecular Plant Pathology, 113, 101574.**

8. **Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016).** ***Using deep learning for image-based plant disease detection*****. Frontiers in Plant Science, 7, 1419**

9. **Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021).** ***Learning transferable visual models from natural language supervision*****. arXiv preprint arXiv:2103.00020.**

10. **Ramesh, M. V., & Vydeki, D. (2020).** ***Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm*****. Information Processing in Agriculture, 7(2), 249–260.**