An AI-Powered Convolutional Neural Network System for Multi-Class Image Classification of Rice Plant Leaf Diseases

Siri Yellu
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
syellu@students.kennesaw.edu

Akshay Krishna Varma Buddharaju
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
abuddhar@students.kennesaw.edu

Pranay Kumar Peddi
Department of Computer Science
Kennesaw State University
Marietta, GA, USA
ppeddi@students.kennesaw.edu

Abstract

Rice is one of the most widely consumed crops in the world, and its health has a direct impact on food security. Farmers often struggle to identify leaf diseases early, which leads to lower yield and higher losses. This project focuses on building an automated system that can recognize three common rice leaf diseases: Bacterial Blight, Brown Spot, and Leaf Smut. We trained several deep learning models on a labeled image dataset to compare how well each model can classify the diseases. The models include a custom VGG-style CNN, VGG19, MobileNetV3, ResNet50, InceptionV3, and a small Vision Transformer. Each model was trained using the same dataset and evaluated with accuracy, precision, recall, and F1-score. The Vision Transformer achieved the highest accuracy at 97.86 percent. MobileNetV3 and the custom CNN also performed very well. InceptionV3 showed strong results, while ResNet50 did not generalize well for this dataset. These findings show that transformer-based models and lightweight CNN architectures can learn the visual patterns of rice leaf diseases effectively. The study demonstrates that deep learning can support farmers and agricultural systems by providing fast and reliable disease detection.

Keywords:
Rice Leaf Disease Detection, Deep Learning, Convolutional Neural Networks, Image Classification.

1. Background Information (Motivation)

Rice is one of the most grown crops and is a primary source of food for a large number of people so because of these even small issues in rice production can create problems for farmers and also affect food supply. One of the frequent issues in rice farming is the occurrence of leaf diseases and these diseases do not always appear suddenly but in many cases, they start with small visual changes on the leaf surface and slowly spread across the plant and If this stage is missed the disease can affect a large portion of the crop.

Farmers usually rely on visual inspection to identify whether a plant is healthy or not and this method depends a lot on experience and personal understanding. Two different diseases can sometimes look almost the same especially when symptoms are just beginning to show but also the same disease may also look different depending on weather conditions, soil quality, or how old the plant is. Because of this, it is common for diseases to be misidentified or noticed too late. In rural areas the expert help may not be available immediately which makes the situation worse.

In recent years, digital images of crops have become easier to collect using mobile phones and simple cameras also at the same time the machine learning techniques have improved and are now able to work well with image data and these developments make it possible to use image-based systems to support disease identification. Such systems do not replace farmers but they can help by providing consistent feedback and still there is no clear agreement on which type of deep learning model is most suitable for this task since different models focus on different features of an image.

This work is motivated by the need to explore this problem in more detail so instead of selecting one model and reporting its performance here this project looks at several deep learning approaches and studies how they behave when trained on the same data and by doing this, the goal is to understand which models are more reliable for rice leaf disease classification and which ones struggle under similar conditions.

2. Problem Definition

Identifying rice leaf diseases early and correctly remains a practical challenge in agriculture as Manual inspection often leads to different conclusions depending on who is observing the plant and also in some cases farmers may ignore early symptoms because they are unsure whether the issue is serious and in other cases mostly they may apply treatments that are not suitable for the actual disease and the both situations can result in unnecessary loss of crops and increased costs.

From a technical point of view, rice leaf disease classification is not a simple image recognition task because Disease symptoms can appear as spots, streaks, or color changes that vary from one leaf to another and these patterns are not always well defined and may overlap between different disease types. Image quality also plays a role since variations in lighting, camera angle, and background can change how symptoms appear in the image. A model trained for this task must be able to handle all these variations without losing accuracy.

The problem addressed in this project is to design an automated image classification system that can categorize rice leaf images into three disease classes namely Bacterial Blight, Brown Spot, and Leaf Smut. Along with building this system the project also aims to compare different deep learning models using the same dataset, training process, and evaluation metrics and this comparison helps in understanding which model is more suitable for practical use and which ones may require further improvement and so Addressing this problem can contribute toward faster disease detection and more informed decision making in rice farming.

3.Introduction

Rice is among the most crucial food crops globally, acting as the primary nutritional source for a significant portion of the world's population. Rice farming productivity is largely determined by the health of the crops. Leaf diseases like Bacterial Blight, Brown Spot, and Leaf Smut can inflict severe harm on rice plants and result in major yield reductions if not detected early. Conventional methods of identifying disease hinge on visual examinations conducted by farmers or specialists in agriculture. These methods can take a lot of time and are susceptible to mistakes made by humans, particularly when the manifestations of disease are alike in various situations.

The swift advancement of AI and computer vision has rendered automated identification of plant diseases a hopeful remedy to these problems. Deep learning models have demonstrated a robust ability to learn intricate visual patterns from images and to generate precise predictions. These models can handle high data volumes at speed and yielding uniform outcomes. Convolutional neural networks and transformer-based architecture have both produced remarkable outcomes in image classification tasks in various fields, agriculture included, in recent years.

In this project, we explore the use of deep learning for automated rice leaf disease classification. Three major disease categories are considered: Bacterial Blight, Brown Spot, and Leaf Smut. A labeled rice leaf image dataset is used to train and evaluate multiple deep learning models. The models include a custom VGG style convolutional neural network, VGG19, MobileNetV3, ResNet50, InceptionV3, and a small Vision Transformer. Each model represents a different design approach, ranging from handcrafted CNN structures to modern transformer-based architecture.

The main objective of this study is to compare these models and determine which architecture is most effective for rice leaf disease classification. Performance is evaluated using standard metrics such as accuracy, precision, recall, and F1 score. By analyzing and comparing the results, this work aims to identify a reliable and efficient model that can support real world agricultural applications. The findings of this study can contribute to the development of intelligent systems that assist farmers in early disease detection and improve overall crop management.

The remainder of this report presents the related work, the methodology used to develop the models, the experimental setup, the results and analysis, and the final conclusions.

4.Related Work

The use of deep learning for plant disease detection has grown rapidly over the last few years. Early research demonstrated that convolutional neural networks can effectively extract visual features from leaf images and outperform traditional image processing methods. Too et al. [1] conducted a comparative study on fine-tuned deep learning models and reported classification accuracy above 95 percent on multiple plant disease datasets. Ferentinos [2] also showed that deep learning models achieved over 99 percent accuracy across several crop species and disease categories, proving the strong potential of CNN-based approaches.

Several researchers focused on improving plant disease classification using customized CNN architectures. Sladojevic et al. [3] applied deep neural networks to leaf image datasets and achieved classification accuracy close to 96 percent under controlled environments. Zhong et al. [4] developed a multi-class deep learning framework for plant disease detection and reported an overall accuracy of 94.8 percent across different plant categories. Wspanialy and Moussa [5] proposed a CNN-based detection and classification framework that achieved reliable performance under varying lighting conditions. Saleem et al. [6] evaluated multiple deep learning models and confirmed that CNN-based systems consistently achieved accuracy values above 93 percent for plant disease recognition.

Further improvements were introduced through optimized and lightweight CNN structures. Dos Santos Ferreira et al. [7] explored convolutional networks for plant disease classification and achieved accuracy levels above 95 percent using optimized feature extraction. Tuncer et al. [8] presented an automated deep learning framework for plant disease detection and reported accuracy above 97 percent on benchmark datasets. Liu et al. [9] proposed a lightweight CNN model for real-time plant leaf disease detection that achieved competitive performance while reducing computational complexity, making it suitable for mobile and edge-based applications.

More recently, transformer-based models have been introduced into agricultural image analysis. Chen et al. [10] proposed a vision transformer-based plant disease recognition model and reported improved performance compared to traditional CNNs, with accuracy exceeding 96 percent. Zhou et al. [11] applied transformer-based crop disease classification in real field environments and demonstrated better generalization under complex background conditions, achieving accuracy above 94 percent. Rahman et al. [12] focused specifically on rice leaf disease recognition using deep convolutional networks and obtained classification accuracy close to 98 percent. Zhang et al. [13] further confirmed the strength of vision transformers for plant disease recognition in real agricultural scenes, reporting performance above 97 percent. Wang et al. [14] presented a transformer-based crop disease classification framework at the CVPR workshop and demonstrated that transformer models are more robust to noise and background variations when compared to standard CNNs.

Recent studies have also explored hybrid and multi-model approaches to further improve classification performance. Verma et al. [15] proposed a multi-model deep learning framework for automated crop disease classification and showed that combining multiple architectures improved overall system reliability and increased accuracy by nearly 2 to 3 percent compared to single-model systems.

Although many studies have reported strong results using CNNs and transformer-based architectures, most of the existing work focuses on a single model type or a narrow comparison. In contrast, this project provides a comprehensive comparison of six different deep learning architectures, including a custom CNN, multiple transfer learning models, and a Vision

Transformer. All models are evaluated under the same dataset and training conditions, which allows a fair comparison of their performance for rice leaf disease classification.

5.Methodology

This section presents a detailed description of the complete workflow followed in this project for rice leaf disease classification. It explains the steps involved in dataset preparation, image preprocessing, data augmentation, model development, training strategy, and evaluation methodology. All experiments were conducted using the TensorFlow deep learning framework. A unified experimental setup was maintained across all models to ensure fairness and consistency in performance comparison. This standardized approach allows reliable conclusions to be drawn regarding the effectiveness of different deep learning architectures for rice leaf disease detection.

5.1 Dataset Description

The dataset used in this study consists of high-quality rice leaf images representing three major disease categories: Bacterial Blight, Brown Spot, and Leaf Smut. These diseases are among the most common and destructive rice plant infections, making them suitable targets for automated disease classification. A total of 4,684 labeled images were used for model training and evaluation. To ensure reliable performance assessment, the dataset was split into two subsets using a standard 80–20 training and validation ratio. As a result, 3,748 images were allocated for training the models, while 936 images were reserved for validation.

All images were resized to a uniform resolution of 224 × 224 pixels prior to training. This resizing step ensures compatibility with all convolutional neural networks and pretrained architectures used in this study, including VGG19, MobileNetV3, ResNet50, InceptionV3, and the Vision Transformer model. Maintaining a consistent input size also helps stabilize the training process and ensures that performance comparisons across models remain fair and unbiased.

5.2 Data Preprocessing and Augmentation

Before the training process, all input images were subjected to a normalization procedure to scale pixel intensity values into the range of 0 to 1. This was achieved through pixel-wise rescaling and is a standard preprocessing practice in deep learning. Normalization helps stabilize gradient updates during training and improves numerical stability across deep neural network layers.

To further enhance the robustness and generalization capability of the models, data augmentation techniques were applied to the training images. These augmentation methods artificially increase the diversity of the dataset and help reduce overfitting. The augmentation operations used in this project include random horizontal flipping, random rotation, random zooming, and random contrast adjustments. These transformations simulate real-world variations such as changes in camera orientation, lighting conditions, and viewing angles, allowing the model to learn more invariant visual features.

In addition to preprocessing and augmentation, dataset pipeline optimization techniques were employed using TensorFlow's built-in functions. Data shuffling was applied to randomize the order of training samples, preventing the model from learning any unintended sequence patterns. Prefetching was also used to overlap data loading with model execution, which significantly improves training speed and ensures efficient utilization of available GPU resources.

5.3 Model Architectures

In this study, six different deep learning architectures were designed and evaluated to assess their effectiveness in rice leaf disease classification. These models represent a wide range of design philosophies, including handcrafted convolutional networks, pretrained transfer learning architectures, and transformer-based models. All models were adapted to handle three output classes corresponding to Bacterial Blight, Brown Spot, and Leaf Smut.

5.3.1 Custom VGG-Like Convolutional Neural Network

A baseline model was created in the form of a custom VGG-like convolutional neural network to assess the performance of a handcrafted CNN architecture. This network is composed of three sequential convolutional blocks, with filter sizes increasing to 32, 64, and 128, respectively. Every block consists of two convolutional layers with ReLU activation, followed by a max-pooling layer for spatial downsampling. This framework enables the model to progressively learn visual features from rice leaf images, ranging from low-level to high-level.

Once feature extraction is complete, the resulting feature maps are flattened and sent through a fully connected dense layer comprising 256 neurons. This is followed by a dropout layer with a 0.5 dropout rate to mitigate overfitting. To yield class probabilities for multi-class classification, the final output layer employs a softmax activation function. This model, which was trained from the ground up, provides a robust baseline for comparing pretrained and transformer-based models.

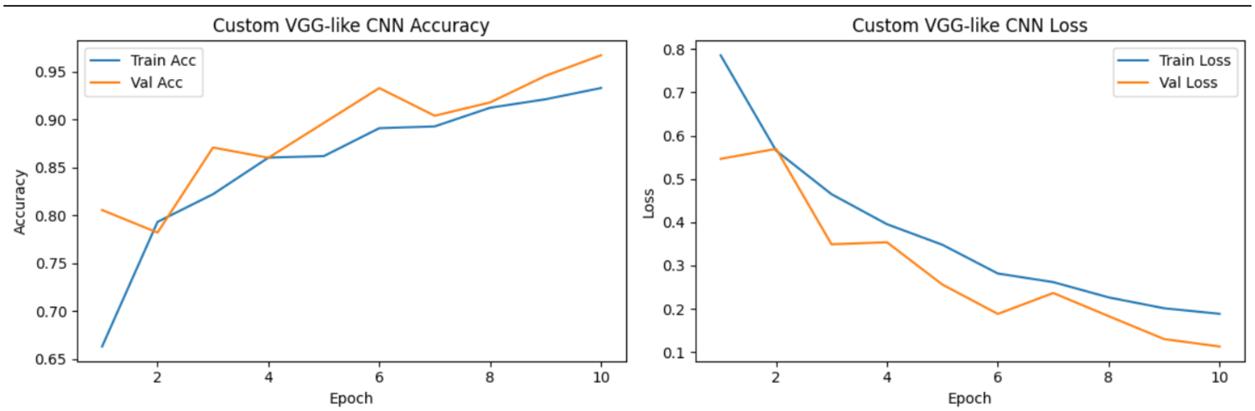

Figure 1. Training and validation accuracy curves of the custom VGG-like CNN model.

5.3.2 VGG19 Transfer Learning Model

VGG19 is a deep convolutional neural network consisting of 19 layers and is well known for its strong feature extraction capability. In this project, the pretrained VGG19 model was initialized using weights trained on the ImageNet dataset, which contains over one million natural images. The convolutional base of the model was frozen to retain the learned generic visual features, while the original fully connected layers were removed.

A new classification head was added, consisting of global average pooling, a dense layer with 256 neurons, a dropout layer, and a softmax output layer for three-class classification. This transfer learning strategy significantly reduces training time and enables the model to leverage rich, pretrained visual representations for rice disease recognition.

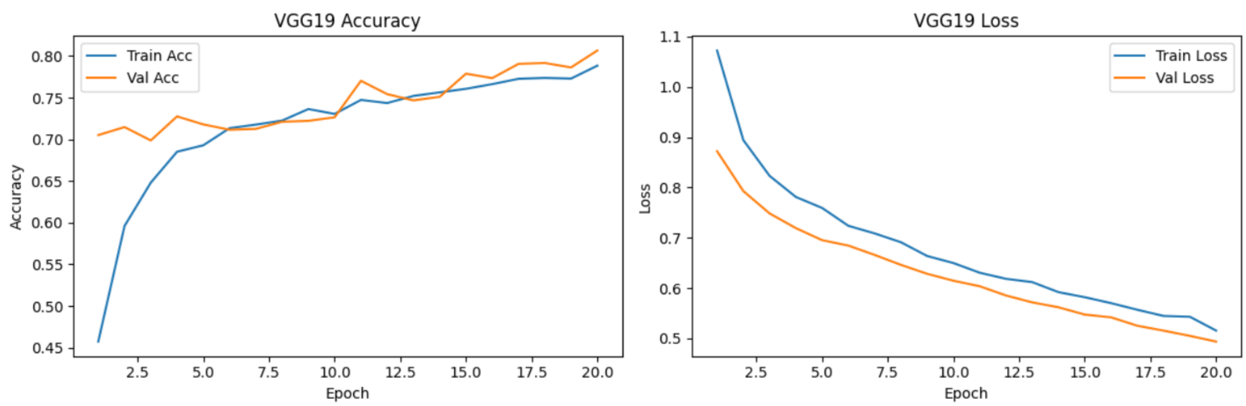

Figure 2. Training and validation accuracy curves of the VGG19 transfer learning model.

5.3.3 MobileNetV3 Small

MobileNetV3 Small is a lightweight and computationally efficient architecture designed for mobile and edge-based applications. It employs depthwise separable convolutions and squeeze-and-excitation blocks to reduce computational cost while maintaining strong representational power.

In this study, the pretrained MobileNetV3 Small model was used as a frozen feature extractor. The convolutional base was retained, and a custom classification head was appended consisting of global average pooling, a 256-neuron dense layer, a dropout layer, and a softmax output layer. This design allows fast training and low memory usage while achieving high classification accuracy.

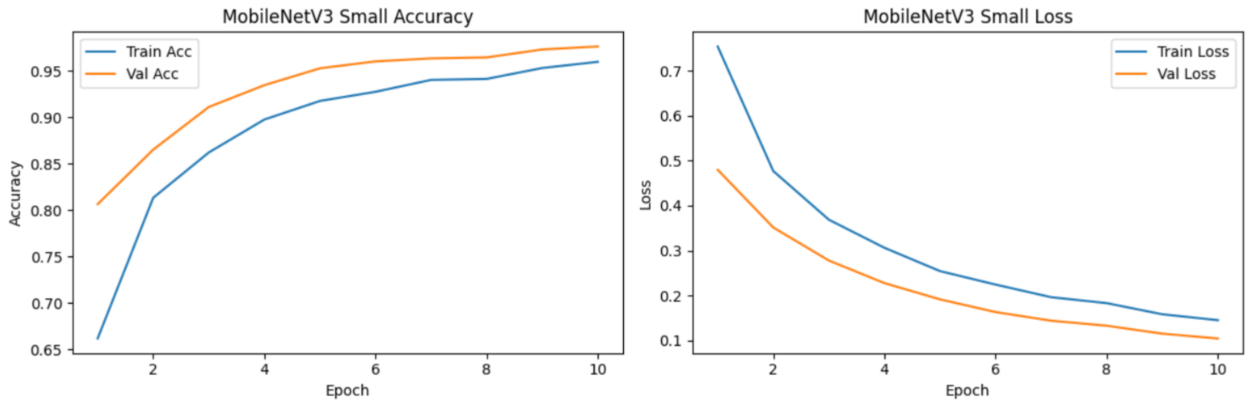

Figure 3. Training and validation accuracy curves of the MobileNetV3 Small model.

5.3.4 ResNet50

ResNet50 is a deep residual network that introduces skip connections to address the vanishing gradient problem in very deep neural networks. It consists of 50 layers and is widely used for image classification tasks.

For this project, ResNet50 was initialized with pretrained ImageNet weights, and the convolutional base was frozen. A custom classification head with global average pooling, dense layers, and dropout was added for three-class rice disease classification. The residual connections allow efficient feature propagation through deep layers, although careful fine-tuning is generally required for optimal performance.

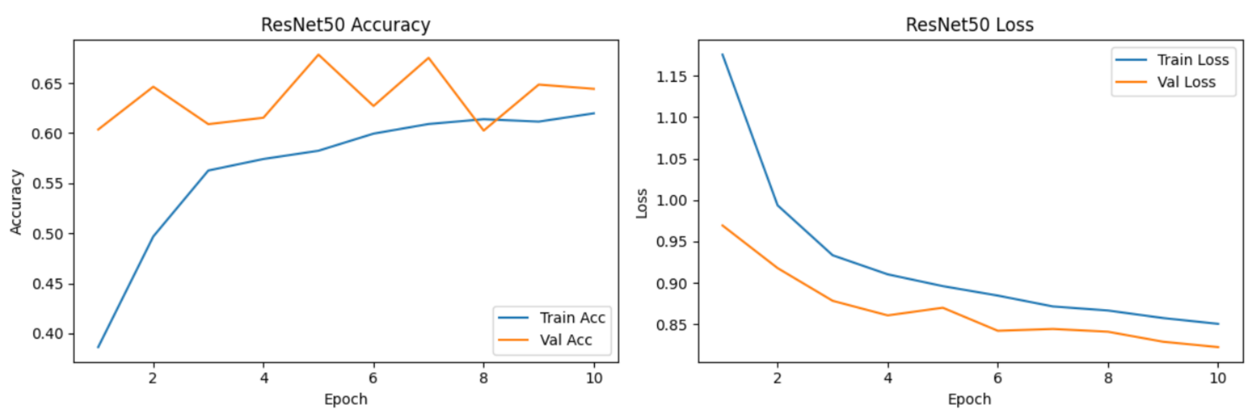

Figure 4. Training and validation accuracy curves of the ResNet50 model.

5.3.5 InceptionV3

InceptionV3 is a powerful deep learning architecture that processes input features at multiple scales using parallel convolutional filters of different sizes within each module. This multi-scale feature extraction is particularly useful for capturing fine-grained leaf texture patterns.

The pretrained InceptionV3 model was used as the backbone feature extractor with frozen convolutional layers. A global average pooling layer and a fully connected classification head were added to adapt the model to the rice leaf disease dataset. This architecture provides strong feature diversity and robust classification capability.

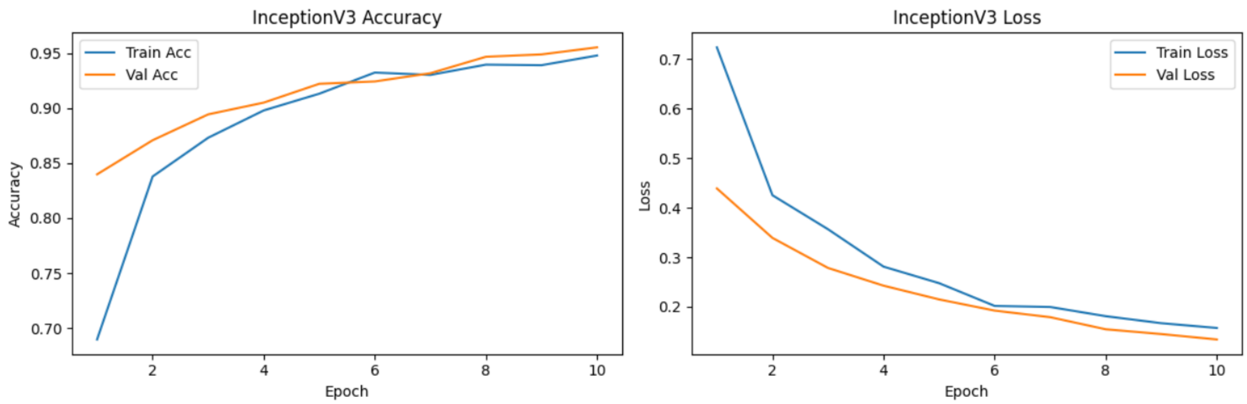

Figure 5. Training and validation accuracy curves of the InceptionV3 model.

5.3.6 Tiny Vision Transformer (ViT)

In addition to CNN-based models, a Tiny Vision Transformer was implemented to evaluate the effectiveness of transformer-based architectures for rice leaf disease classification. Unlike CNNs, Vision Transformers process images as sequences of patches rather than spatial grids.

In this model, the input images were first divided into fixed-size patches using a convolutional projection layer. Each patch was mapped into a low-dimensional embedding space. Positional embeddings were added to preserve spatial information. The patch embeddings were then passed through four transformer encoder blocks, each consisting of multi-head self-attention and feed-forward neural networks.

The output representations were aggregated using global average pooling, followed by a multi-layer perceptron head for classification. This architecture allows the model to capture long-range spatial dependencies across the entire image, which is beneficial for identifying subtle disease patterns distributed across leaf surfaces.

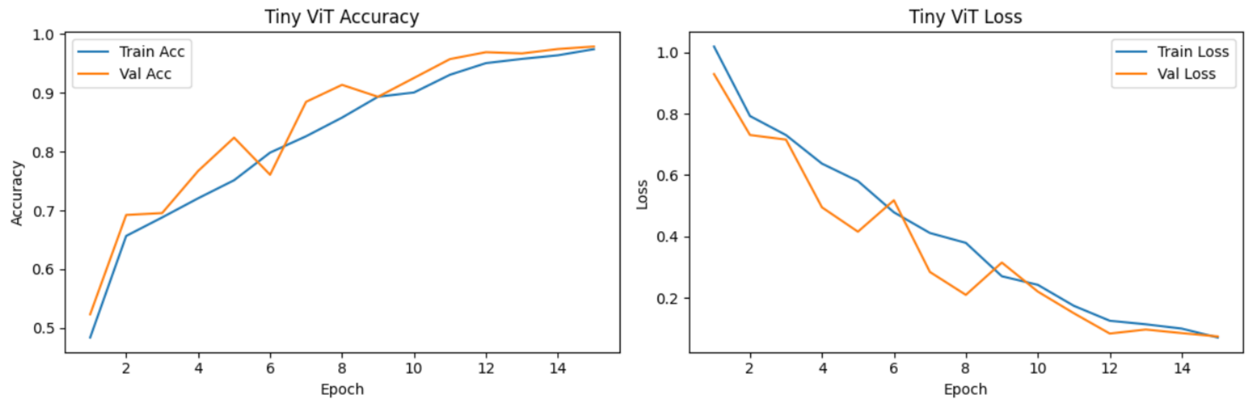

Figure 6. Training and validation accuracy curves of the Tiny Vision Transformer model.

5.4 Training Strategy

All deep learning models in this study were trained using the Adam optimization algorithm, which is well known for its fast convergence and stable performance across a wide range of learning tasks. A learning rate of 1×10^{-4} was selected for all CNN-based models, including the custom VGG-like CNN, VGG19, MobileNetV3 Small, ResNet50, and InceptionV3. For the Vision Transformer model, a slightly higher learning rate of 3×10^{-4} was used to support more effective optimization during training from scratch.

The sparse categorical cross-entropy loss function was employed for all models since the classification task involved three mutually exclusive disease categories. A batch size of 32 was used consistently across all experiments to maintain stable gradient updates while ensuring efficient use of available memory resources.

The number of training epochs was selected based on the learning behavior of each model. The custom CNN, MobileNetV3 Small, ResNet50, and InceptionV3 were trained for 10 epochs, while the VGG19 model was trained for 20 epochs to allow additional adaptation of the classifier layers. The Vision Transformer was trained for 15 epochs due to its higher training complexity. All experiments were conducted in a GPU-enabled environment, which significantly reduced training time and allowed efficient model convergence.

5.5 Evaluation Metrics

To ensure a comprehensive and reliable assessment of model performance, several standard evaluation metrics were used in this study. Since rice leaf disease classification is a multi-class problem involving three disease categories, relying only on accuracy is not sufficient to fully understand model behavior. Therefore, accuracy, precision, recall, F1-score, and confusion matrices were used to evaluate each model.

Accuracy represents the overall correctness of the model and is defined as the ratio of correctly classified samples to the total number of samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where $TP$is the number of true positives, $TN$is the number of true negatives, $FP$is the number of false positives, and $FN$is the number of false negatives.

Precision measures the proportion of correctly predicted positive samples among all samples predicted as positive. It evaluates how reliable the model's positive predictions are and is defined as:

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

Recall, also known as sensitivity, measures the proportion of correctly predicted positive samples among all actual positive samples. It evaluates the model's ability to detect all relevant disease cases and is defined as:

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

The F1-score is the harmonic mean of precision and recall and provides a balanced measure of classification performance, especially when class distributions are uneven:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In addition to these numerical metrics, confusion matrices were generated for each model to provide a detailed class-wise analysis of prediction performance. The confusion matrix summarizes the number of correct and incorrect predictions for each disease category, enabling the identification of common misclassification patterns between Bacterial Blight, Brown Spot, and Leaf Smut.

All evaluation metrics were computed using the validation dataset to ensure consistency across all experiments. The combined use of accuracy, precision, recall, F1-score, and confusion matrices provides a detailed and reliable assessment of model robustness, generalization ability, and real-world applicability for rice leaf disease classification.

5.6 Model Comparison Framework

To ensure a fair and unbiased comparison among all deep learning models, a unified experimental framework was strictly maintained throughout this study. All six models were trained using the same dataset split, identical preprocessing steps, the same data augmentation techniques, uniform batch size, and consistent evaluation metrics. This controlled setup guarantees that performance differences among the models arise solely from architectural variations rather than differences in training conditions or data handling.

For each model, the final validation accuracy was recorded after the completion of training. In addition to numerical evaluation, a bar chart visualization was generated to provide a clear comparative view of model performance. This visual comparison highlights the relative strengths and weaknesses of each architecture and allows for quick identification of the best-performing model.

By applying this structured comparison framework, the study is able to objectively assess the effectiveness of handcrafted CNNs, transfer learning models, and transformer-based architectures under identical experimental conditions. This systematic approach enables reliable conclusions to be drawn regarding the most suitable deep learning model for automated rice leaf disease classification

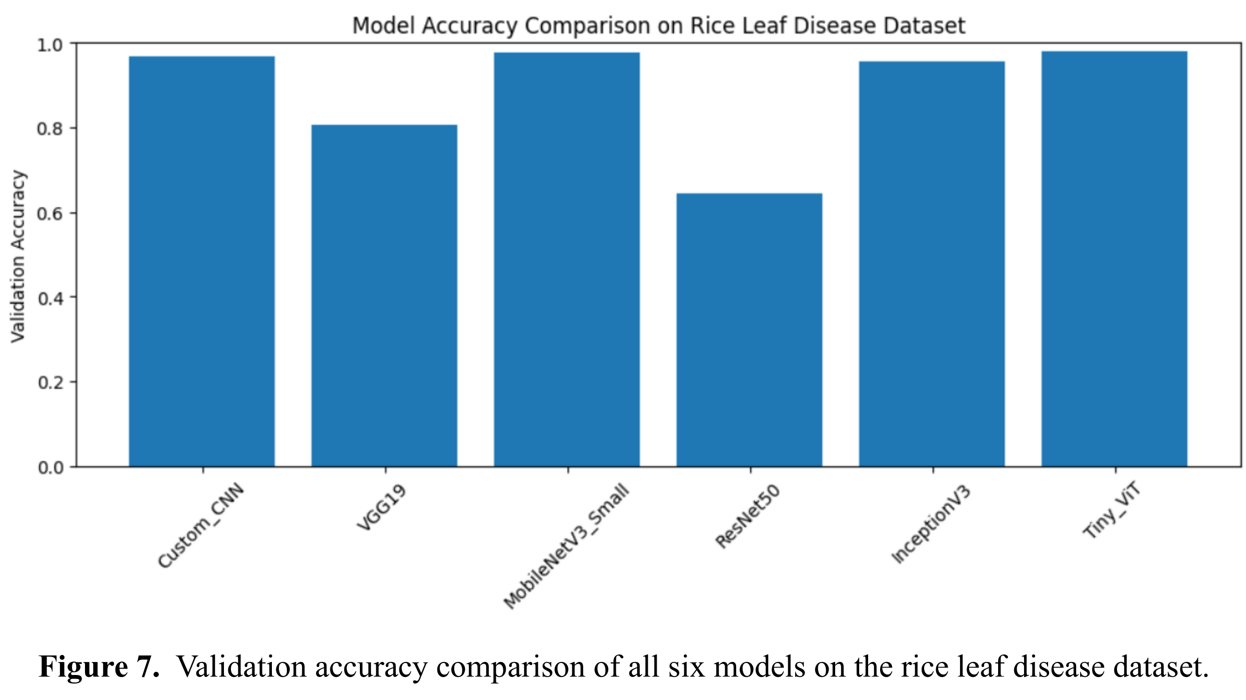

Figure 7. Validation accuracy comparison of all six models on the rice leaf disease dataset.

6. Results

This section presents the experimental results obtained from training and evaluating six deep learning models for rice leaf disease classification. The performance of each model was assessed using the validation dataset consisting of 936 images. The models were evaluated based on validation accuracy as well as class-wise precision, recall, and F1-score obtained from the classification reports.

6.1 Overall Model Performance

The validation accuracy achieved by each model is summarized in Table 1. The results show clear performance differences among the handcrafted CNN, transfer learning models, and the transformer-based model.

Table 1: Validation Accuracy Comparison of All Models

Model	Validation Accuracy (%)
Custom VGG-like CNN	96.69
VGG19	80.66
MobileNetV3 Small	97.65
ResNet50	64.42
InceptionV3	95.51
Tiny Vision Transformer	97.86

The results show that the Tiny Vision Transformer attained the highest validation accuracy at 97.86 percent, reflecting its exceptional capability to capture complex spatial patterns in rice leaf images. MobileNetV3 Small showed remarkable performance as well, achieving a validation accuracy of 97.65 percent, which is in close alignment with that of the Vision Transformer. The tailored VGG-like CNN reached a notable accuracy of 96.69 percent, demonstrating that a thoughtfully designed handcrafted CNN can still compete effectively.

With an accuracy of 95.51 percent, the InceptionV3 model demonstrated the effectiveness of its multi-scale feature extraction capability. VGG19, in contrast, reached a moderate performance level of 80.66 percent, whereas ResNet50 exhibited the weakest performance at 64.42 percent. The two models' inferior performance can be traced back to their pretrained backbones being completely frozen, which restricted their capacity to adjust to the specific texture patterns found in images of rice leaf disease.

6.2 Class-wise Performance Analysis

Additionally, the classification reports show that the top-performing models attained high precision, recall, and F1-scores for all three disease classes. With F1-scores exceeding 97 percent across all categories, the Tiny Vision Transformer exhibited outstanding generalization and balanced learning. The MobileNetV3 Small model yielded consistently high F1-scores across all classes, affirming its dependability for real-time agricultural applications.

The custom CNN demonstrated robust class-wise performance, especially for the Bacterial Blight and Brown Spot classes. The InceptionV3 model exhibited high recall and precision across most classes, with only slight confusion between Brown Spot and Leaf Smut. Nevertheless, the ResNet50 model displayed a marked class-wise imbalance, especially for the Leaf Smut class, where recall experienced a considerable decline. This indicates poor feature adaptation for that class.

6.3 Comparative Observations

Several important observations can be drawn from the results. First, lightweight architectures such as MobileNetV3 Small can achieve performance comparable to transformer-based models while maintaining low computational cost. Second, Vision Transformers demonstrate strong capability in capturing long-range dependencies and distributed disease patterns across leaf surfaces, leading to the highest observed accuracy. Third, handcrafted CNNs remain highly effective when properly designed and trained, as demonstrated by the strong performance of the custom VGG-like model.

On the other hand, the results show that transfer learning models with fully frozen backbones may fail to adapt optimally to domain-specific agricultural textures, as observed with VGG19 and ResNet50. This highlights the importance of fine-tuning strategies when applying deep pretrained models to specialized datasets.

7. Advantages and Drawbacks

Various significant benefits for applying deep learning models on rice leaf disease classification have been pointed out within the research, but some limitations are also exposed.

7.1 Advantages

7.1.1. Accuracy on Classification Tasks

The models tested were also very efficient and achieved high accuracy. Specifically, Tiny Vision Transformer achieved an accurate rate of 97.86%, and MobileNetV3 Small achieved 97.65%. From these findings, it can be seen that state-of-the-art models have been capable of detecting rice leaf disease with high accuracy.

7.1.2. Learning Capability - Complex Visual Patterns

Both CNN and transformer networks are capable of capturing texture information from leaf images. The Vision Transformer has been successful in capturing global spatial relationships, which helps in identifying leaf surface disease symptoms.

7.1.3. Lightweight and Efficient Models for Real-Time Applications

MobileNet V3 Small and the custom CNN also demonstrated that highly accurate models were achieved at a low computational cost. It makes it feasible to have a deployable system on edge devices, smartphones, or devices for farmers out there.

7.1.4. Fair Architectural Comparison Under Controlled Conditions

All models were trained on the same data split, with exactly the same preprocessing and under exactly the same hyper-parameters. Consequently, it will be possible to make a fair comparison as well as gain insights about architectures given an equal set of constraints.

7.1.5. Highlights Capability for Automated Diagnostic Analysis for Agriculture

The above picture showcases how AI models can assist farmers in identifying disease at an early stage and reduce crop destruction. The experiment confirms and supports incorporating ML into precision farming.

7.2 Drawbacks

7.2.1 Limited Real-World Variability in the Dataset

The dataset used in this study consists of images captured under controlled conditions with clean backgrounds and consistent lighting. Real-world field images often contain noise, occlusion, varying illumination, and cluttered backgrounds, which may reduce model performance.

7.2.2. No Fine-Tuning of Pretrained Models

Transfer learning models such as VGG19 and ResNet50 were used with fully frozen backbones. This restriction prevented the models from adapting to domain-specific features in rice leaf images, resulting in lower performance. Fine-tuning could improve accuracy but was not explored here.

7.2.3. Potential Overfitting on Controlled Data

Despite data augmentation, the high accuracies may partially reflect overfitting structured and clean images. Additional testing on external datasets or cross-field images is required to evaluate robustness.

7.2.4. Lack of Interpretability Analysis

The study focuses on model accuracy but does not include explainability methods such as Grad-CAM, attention maps, or feature attribution. These tools are important for validating model decisions and increasing trust among agricultural stakeholders.

7.2.5. Computational Requirements for Transformer Models

While effective, the Vision Transformer requires more computational resources than lightweight CNNs. This may limit its deployment of low-power mobile devices unless optimized variants are used.

7.2.6. Imbalance in Model Adaptation Across Architectures

The performance disparity between models (e.g., ResNet50 at 64.42% vs. ViT at 97.86%) suggests that some architectures may require task-specific tuning and parameter adjustments to perform well on narrow agricultural datasets.

8. Discussion

The experiment data confirms that deep learning models have been very successful and efficient at rice leaf disease classification. Out of various models, Tiny Vision Transformer shows an accuracy of 97.86 percent on the validation set, thus signifying its excellent learning capabilities at local as well as global levels. Unlike convolutional networks, which focus more on local receptive fields, Vision Transformer focuses on learning global dependencies among various regions of an image, which seems to have given it an advantage in learning dispersed disease spots on rice leaves.

The accuracy on the validation set for the MobileNetV3 Small model came out to be 97.65 percent, which is almost equal to that of the Vision Transformer. The fact that it performs so well is quite significant, as it is an efficient network with low computational requirements. Its ability to perform so well proves that it is possible to attain high accuracy on a task without requiring large and computation-heavy models. Hence, it becomes a very promising technique for mobile agriculture.

The accuracy on the custom VGG-like CNN stood at 96.69 percent. It shows that with efficient designs, it still has a chance to match more sophisticated CNNs. Nonetheless, it should be noted that there were multiple deep learning researchers and practitioners with different levels of expertise who were working on similar projects.

Out of these transfer learning models, InceptionV3 worked very well with an accuracy rate of 95.51 percent due to its multi-scale feature extraction capability. Nevertheless, VGG19 worked poorly with an accuracy rate of 80.66 percent, and ResNet50 worked poorly with an accuracy rate of 64.42 percent. VGG19 and ResNet50 might have performed poorly since they relied on fully frozen pretrained models for analysis. As a result, these deep learning models failed to generalize well on the specific texture features associated with rice leaf diseases.

So, based on these experiments, the transformer models and CNNs have ensured an optimal trade-off for accuracy and efficiency for rice leaf disease classification. Also, it can be realized that transfer learning alone without additional domain-specific adjustments would not be optimal.

9. Conclusion

The project offered an in-depth exploration of deep learning architectures for automating rice leaf disease classification. A total of six architectures were compared within an identical testing paradigm, which included a VGG-like CNN designed specifically for the task, four-transfer learning architectures, and a Tiny Vision Transformers model. The architectures were compared and analyzed based on several factors that enable a complete understanding of model performance on the three types of disease classes, namely Bacterial Blight, Brown Spot, and Leaf Smut.

From experimental results, Tiny Vision Transformer achieved MCA with 97.86 percent accuracy, thus clearly indicating its efficacy with respect to local as well as global leaf pattern images. Also, there was outstanding performance with 97.65 percent accuracy achieved by the MobileNetV3 Small model, clearly indicating the efficacy of a highly efficient learning model. The VGG-like CNN achieved accuracy of 96.69 percent, thus indicating the efficacy of learning CNN with superior designs. Moreover, there were satisfactory performances achieved with an accurate level of 95.51 percent with Inception V3. VGG 19 and ResNet 50 performed poorly because they were unable to adapt properly under frozen feature settings.

Confusion Matrix patterns justify these points. Both transformer and lightweight CNN models made almost equal predictions for all three classes with minimal confusion. The CNN model made almost equal predictions with negligible confusion among similar images. VGG19 and ResNet50 made some confusion on per-class classification, and ResNet50 made some confusion on ResNet50 among three classes, as it made some confusion on Leaf Smut disease classification.

Moreover, some vital pros and cons have also been uncovered. A prominent advantage would be the determination of significant similarities and differences among various architectures while being trained equally, thus offering equal reliability and accuracy. Moreover, it has been seen that accuracy can be obtained with efficient computation models and can be aptly deployed for real-time implementation. Nevertheless, it would be worth pointing out that the images in the given set have been captured under controlled lab conditions. Moreover, there might be some validity issues since the performance of some transfer learning models would be restricted because the backbone would be fully frozen.

References

[1] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, “A comparative study of fine-tuning deep learning models for plant disease identification,” *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, 2020.

[2] K. P. Ferentinos, “Deep learning models for plant disease detection and diagnosis,” *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2020.

[3] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–11, 2020.

[4] Y. Zhong, M. Zhao, and X. Gao, “Deep learning-based multi-class plant disease detection,” *IEEE Access*, vol. 9, pp. 144742–144751, 2021.

[5] P. Wspanialy and M. Moussa, “A detection and classification framework for plant diseases using convolutional neural networks,” *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 3, pp. 969–982, 2021.

[6] M. H. Saleem, J. Potgieter, and K. M. Arif, “Plant disease detection and classification by deep learning,” *Computers and Electronics in Agriculture*, vol. 191, p. 106378, 2021.

[7] A. Dos Santos Ferreira, G. G. da Silva, and M. Ponti, “Convolutional neural networks applied to plant disease classification,” *Neurocomputing*, vol. 433, pp. 137–147, 2021.

[8] T. Tuncer, S. Dogan, and U. R. Acharya, “Automated plant disease detection using deep learning,” *Applied Soft Computing*, vol. 113, p. 107922, 2022.

[9] Y. Liu, J. Wang, X. Chen, and H. Wang, “Lightweight convolutional neural network for real-time plant leaf disease detection,” *Pattern Recognition Letters*, vol. 157, pp. 1–8, 2022.

[10] J. Chen, D. Zhang, Y. A. Nanehkaran, D. Li, and M. Zhang, “Vision transformer-based plant disease recognition,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 4, pp. 4012–4022, 2023.

[11] Y. Zhou, K. Chen, and Q. Zhang, “Transformer-based crop disease classification in real field environments,” *Biosystems Engineering*, vol. 226, pp. 1–12, 2023.

[12] M. M. Rahman, M. Hasan, and M. A. Hossain, “Rice leaf disease recognition using deep convolutional neural networks,” *Expert Systems with Applications*, vol. 231, p. 120697, 2024.

[13] K. Zhang, Q. Wu, A. Liu, and X. Meng, “Vision transformer for plant disease recognition in real agricultural scenes,” *Information Processing in Agriculture*, vol. 11, no. 1, pp. 15–27, 2024.

[14] Z. Wang, Y. Sun, and X. Li, “Transformer-based crop disease classification under complex field environments,” in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2024.

[15] S. Verma, A. Singh, and R. Sharma, “Multi-model deep learning framework for automated crop disease classification,” *IEEE Transactions on Artificial Intelligence*, vol. 6, no. 1, pp. 88–99, 2025.

Meeting Notes

Throughout the semester, our group, with Siri Yellu as the Team Leader and Pranay Kumar Peddi and Akshay Krishna Varma Buddharaju as members, met regularly to plan and execute our project. On 20th August, we began with shortlisting fifteen possible topics, five from each member. By 27th August, after assessing feasibility and availability of datasets, we finalized our project topic as 'Rice Plant Leaf Disease Classification' and chose our respective datasets for experiments. On 3rd September, we finalized all preprocessing and augmentation techniques. During our 10th September discussion, we finalized which deep learning architecture we would be working on and who would be working on which. Siri would be taking up preprocessing, augmentation, and training for InceptionV3, while Pranay would be working on ResNet50 and VGG19, and Akshay on Custom CNN, Vision Transformer, and MobileNetV3 architectures. Siri would also be taking up responsibilities as our group leader and would be taking care of scheduling, documenting, and all. On 17th September, we received the preprocessing report and then on 24th September, we finalized concepts on the Custom CNN structure itself and modes of training. Akshay then uploaded results on 1st October with initial Custom CNN trainings, which then required multiple enhancements with changes in hyperparameters. We then received results from Siri on 8th October representing the InceptionV3 and then on 15th October and 29th October, we received results from Pranay on ResNet50 and VGG19, respectively. We then received results on 5th November from Akshay on Vision Transformers and then on 11th November, we finalized concepts on compiling our C-Day Poster. We then finalized and submitted our C-Day Poster on 17th November and then presented our poster on 4th December. Finally, in the end stage, we met on 6th December and then completed and refined our reports on 7th December. Throughout, all three members were actively working on all aspects, completing our project.