An efficient Deep Learning approach to Detect Plant Diseases using Image Data

by

AHMAD ANIS
22101254
BABLI SAHA
21241002
MAISA TARANNUM SRIZEE
22101509
SIKDER MARBIN RIME
22301288
FOYZUNNESA SWARNA
19301220

A thesis submitted to the Department of Computer Science and Engineering in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
October 2025

Declaration

It is hereby declared that:

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where appropriately cited.

3. The thesis has not been submitted for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signatures:

AHMAD ANIS
22101254

BABLI SAHA
21241002

MAISA TARANNUM SRIZEE
22101509

SIKDER MARBIN RIME
22301288

FOYZUNNESA SWARNA
19301220

Approval

The thesis titled **"An efficient Deep Learning approach to Detect Plant Diseases using Image Data"** submitted by:

1. Ahmad Anis (22101254)
2. Babli Saha (21241002)
3. Maisa Tarannum Srizee (22101509)
4. Sikder Marbin Rime (22301288)
5. Foyzunnesa Swarna (19301220)

of Summer, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of **B.Sc. in Computer Science and Engineering** on 31 January 2026.

Examining Committee:

Supervisor:
(Member)

Dr. Ashraful Alam
Associate Professor
Department of Computer Science and Engineering
BRAC University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Plant diseases drastically impact on the yield and quality of crops and this is a great challenge to food security and sustainable agriculture in the world and more so in countries such as Bangladesh [18]. There is need to therefore ensure the early and accurate diagnosis of plant diseases to enhance productivity and safeguard food supplies. Plant disease classification based on images has been highly successful with the swift development of deep learning, especially Convolutional Neural Networks (CNNs).In this thesis, a lightweight hybrid CNN-Transformer model is suggested as a classification of rice leaf disease. A number of pretrained CNN models, such as ResNet50, VGG16, InceptionV3, MobileNet V3, EfficientNetB0, and DenseNet121 are tested with transfer learning to determine the baseline performance. The dataset used in the study is the 23, 265 images of rice leaf in 10 different classes including healthy and diseased.To (hopefully) bypass the shortcomings of single-architecture models, the hybrid model proposed replaces local feature extraction with EfficientNetB0 and global contextual learning with TinyViT as well as a lightweight version of MobileNet. The experimental findings indicate that the hybrid method is better than single CNN and independent Transformer models as it has an accuracy of 98.04%with high precision, recall, and F1-score and low costs of computation. The results demonstrate the usefulness of lightweight CNN-Transformer models in real-time and resource-limited agricultural disease prediction and farm health crop management.

Keywords: Plant Disease Detection, Tiny ViT (Tiny Vision Transformer), Deep Learning, Hybrid Model, EfficientNetB0, small MobileNetV3, CNNs (Convolution Neural Network), Swin-Tiny

Acknowledgement

begin with, we would like to thank Almighty Allah and the fact that we managed to finish this thesis. We earnestly wish to acknowledge our supervisor, **Dr. Ashraful Alam**, Assistant Professor, Department of Computer Science and Engineering, BRAC University, who patiently mentored and guided us, gave prompt suggestions and supported us with his unrelenting encouragement during our research and writing.

We would like to give special credit to **Shammo Biswas**, Research Assistant, BRAC University, who has been of great help in the preparation of this paper through the continued technical support and insightful contributions.

We take this opportunity to recognize with gratitude **Dr. Md Golam Rabiul Alam** who is a Professor at BRAC University, and helped in taking care of the hardware resources required to carry out our experiments.

Lastly, we owe our parents a lot of our achievements through their encouragement and support in our dreams. Besides, we owe ourselves the credit of having done this research paper.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The impact of the prevalence of plant diseases has a deeper impact on agricultural productivity and food security in the world, mainly in terms of decrease in quality and yield. Early diagnosis of these diseases in many agrarian environments, especially in economically underprivileged countries, is an extremely urgent issue, and it is primarily due to its dependence on human behavior and professional skills. The current diagnostic technologies are traditionally lengthy, subjective and even impossible to conduct in case of prolonged or continuous observation. As a result, there is often a high cost to agriculture imposed by the absence of quick disease diagnostics, which often leads to a considerable loss of economic resources and compromises the sustainability of agricultural processes.

The latest developments in computer vision and deep learning have made image-based disease detection one of the feasible automated methodologies. Convolutional neural networks (CNNs) are always superior in the ability to extract features peculiar to leaf images and, as a result, they find extensive application in modern studies to classify plant diseases [21]. Pre-trained deep-learning networks (ResNet50, VGG16, MobileNetV3, DenseNet121, InceptionV3, and EfficientNetB0) provide astonishingly good results when comparative analysis of datasets of plant diseases is done using transfer-learning strategies.

Because of such advancements, most of the modern-day algorithms to detect plant diseases are trained and tested in laboratory environments using large and comprehensive datasets. In practice, in a real agricultural scenario, images may have an important amount of color variation, background noise, shadows, and a low amount of data. This leads to a significant decrease in performance of many deep-learning models and thus reduces their usefulness [22]. Methods that are aimed at achieving high accuracy are usually computationally-intensive that are used in edge-based agricultural methods.

Past methodologies have mainly used single-architecture models hence limiting the ability to have heterogeneous feature representations. CNN-based methods have been shown to be useful in detecting micro -colour differences and textural features, but they do not well represent long-term dependencies and world context in leaf images. This weakness becomes especially relevant when one has to discriminate between visually similar pathological cases or a preliminary manifestation of a disease, which requires a full-fledged structural input [23]. The recent studies have proposed transformer-based architectures, such as the Vision Transformer (ViT), which can model global interactions in images [24]. Transformer models are computationally expensive however and require large datasets, which makes them not a viable option in resource-limited agricultural environments. The limitations are reduced by lightweight versions, including the Tiny Vision Transformer (Tiny ViT), which is less explored on the task of plant disease classification.

These problems indicate that there is a critical gap in the literature related to the development of efficient and sustainable deep-learning algorithms, which is where accuracy, flexibility, functionality, and computational efficiency should be gained simultaneously. The classical supervised learning algorithms and monolithic model architectures are not well suited to overcome these challenges in the context of limited data and limited computational resources. As a result, an interesting demand to create hybrid deep-learning approaches that combine complimentary architectural elements to learn features with greater effectiveness and at the same time, be practically deployable, exists.

1.2 Motivation

The main aim of such a project is to come up with a more efficient and correct automated system of detecting plant diseases, which can be implemented in the working agricultural environment. Since manual disease diagnosis has limitations, and the need to discover more intelligent solutions related to agriculture is growing, there is a rising need to find reliable and computationally efficient deep learning models. Another reason that is highly persuasive is based on the drawbacks of the traditional deep learning algorithms, which often focus on predictive power at the cost of the model complexity. Mobile and edge devices are used in precision agriculture with large-scale models being hard to implement.

This means that an effective deployment cannot be done without the creation of a light but efficient model.Moreover, the hybrid systems that combine the convolutional neural network-based feature extraction and transformer-based global-context modeling are not fully explored in the field of plant disease diagnostics. The current research paper aims at establishing whether these hybrid systems can improve the classification performance without compromising efficiency, particularly with realistic experiment conditions.

1.3 Problem Statement

Diseases in plants reduce the quantity and quality of agricultural products, which create significant threats to the world food supply. Advanced and fast recognition of plant diseases based on leaf images has become a suggestion among research topics in precision agriculture. The recent developments in the field of deep learning, specifically, convolutional neural networks (CNNs) have provided promising results in terms of detecting plant diseases, but there are still a number of inherent constraints. Traditional CNN-like models are good at retrieving fine-grained spatial details, such as epidermal texture, chromatic variations and disease presentation. Nevertheless, they are limited in their field of receptivity to detect long-range dependencies and global spatial patterns characteristic of the leaf images [23]. It is a limiting factor especially in isolating phenotypically indistinguishable illnesses or detecting an illness at an early stage that needs delicate global indicators. The resulting issues include the correct identification of various diseases of concurrent plants and the practicality of solutions under real-life conditions, including the mobile performance in resource-limited areas.

A large number of existing frameworks are based on using traditional classifiers such as KAN or fully connected layers which might restrict the ability of the model to effectively exploit nonlinear interaction between features that are acquired via deep feature extractors. More developed classifiers, like Kolmogorov-Arnold Networks (KAN), are capable of helping to make the best decisions by creating bigger nonlinear models. Nevertheless, they have primarily been used in detecting diseases in plants. Transformer-based designs, including Vision Transformers (ViT), are now developed with the purpose of addressing the weaknesses of object recognition in vision tasks. The models are perfect in modeling the global connections through internal attention mechanisms and multilevel structures of features. However, straight Transformer-based methods usually require extensive data and numerous computational resources, which is inappropriate when farming regions are data-limited and when they must run constantly [2]. Complex scaling algorithms are used to ensure that EfficientNet and other efficient CNN structures use a balance between accuracy and processing efficiency. Even when the EfficientNet-based models reduce the number of parameters and the cost of predictions, they are not effective at reflecting global conceptual relationships, which restricts their ability to survive in complex disease classification conditions.

The key issue addressed by this study is that the absence of an efficient deep learning architecture is how efficacious CNN-based feature fusion training is merged with Transformer-based overall contextual model construction, like those of EfficientNetB0, Tiny Vision Transformer (tiny Vit) designs, that can execute high accuracy, faithful feature representation, and cost-effective computation to determine plant diseases through the use of images. This is a critical challenge to overcome in order to develop flexible, precise, and easy-to-access detection of plant diseases to be applied in real agricultural settings.

1.3.1 Task Difficulty

It is intrinsically difficult to detect multi-class rice leaf diseases due to a number of reasons. The visual symptoms of many diseases are subtle and overlapping, particularly in their early stages of infection, and it is challenging to differentiate between them with the help of traditional models. The environmental differences, including light variations, ambient noise, and the positioning of the leaf, also make feature extraction difficult. Low labeled dataset availability elevates the chances of overfitting, especially when it has to deal with rare classes of diseases. Also, it is necessary to balance accuracy in classification and efficiency when deploying deep learning models in resource-constrained (including edge devices or mobile platforms) settings. These limitations demonstrate that it is necessary to have a hybrid CNN-Transformer network with complex classifiers such as Kolmogorov-Arnold Networks (KAN) to capture both local and global contextual-specific features and remain feasible in the real world and in agriculture.

1.4 Research Objectives

The proposed system will deliver fast, precise, and scalable diagnosis of plant diseases based on the application of advanced CNN structures, including EfficientNetB0, ResNet50, MobileNetV3, DenseNet121, InceptionV3, and VGG16, as well as effective preprocessing and augmentation strategies [13]. It also has low computational complexity, which is appropriate in a real-life agricultural setting. This method will help farmers make sound decisions that will minimize the losses of crops and increase their yield, enhancing the speed and accuracy of detection.

The major aims of this study are in order to assess the efficiency of commonly used deep-learning models to detect plant diseases with the help of image data, such as ResNet50, MobileNetV3, VGG16, DenseNet121, InceptionV3, and EfficientNetB0. To learn regional image patterns and global context-related information, we will design a deep learning model, which is a hybrid consisting of similar representations of features of EfficientNetB0, Tiny Vision Transformer (Tiny ViT), and a small mobile architecture, MobileNet.

The capability of feature fusion techniques to enhance the accuracy of illness diagnosis with the help of multiple models, deep visual representations of features. To explore the application of Kolmogorov-Arnold Networks (KAN) to act as a sophisticated classifier by contrasting its performance with the conventional layers of classifier of data. To generate a perfect compromise among classification accuracy, reliability, and computing efficiency, permitting it to be used with limited resource applications.

To explore the ability of feature fusion methods for boosting illness diagnosis accuracy by merging multiple models' deep visual representations of features. To investigate the use of Kolmogorov-Arnold Networks (KAN) to serve as an advanced classifier by comparing its effectiveness to traditional classifier layers of data. To produce an ideal balance of classification accuracy, reliability, and computing effectiveness, allowing for effective deployment on limited resource applications.

1.5 Methodology in Brief

This paper uses a proven deep learning-oriented methodology to develop a useful and high-quality disease detection system using leaf visual information. Similar data sets of plant leaves are available widely as a measure to maintain diversity among plant varieties and categories of diseases. To enhance stability to changes in brightness, placement, and dimension, a comprehensive preparation procedure is employed and involves image downsizing, image normalization, background removal, noise, and extensive data augmentation processes.

A wide range of popular pre-trained convolutional neural network structures, including VGG16, ResNet50, MobileNetV3, DenseNet121, InceptionV3, and EfficientNetB0, are optimized with the help of transfer learning. To be given a fair evaluation, these models are evaluated using similar learning setups. Efficiency is calculated using the typical classification parameters including accuracy, precision, recall, F1-score, and confusion matrix estimation, that can give a full picture of class-specific forecasting performance.

We propose a hybrid deep learning system in order to address the basic shortcomings of single-model-based approaches to the complex variations of illness. EfficientNetB0, Tiny ViT, and a small MobileNet model are used as they are beneficial to the system because of the advantages of spatial and conceptual features. Thus, there are deep representations of features extracted from each of the cores that are combined with feature fusion processes to form an overall and more specific set of features, therefore improving disease separation.

Similarly, it is the combination of features that is defined by both classic machine learning classifiers and Kolmogorov-Arnold Network (KAN) to evaluate their ability to be able to capture complex nonlinear boundaries in decision-making. The performance of KAN has been compared to normal classifiers with a preference on classification accuracy and flexibility.

The proposed structure is strictly evaluated using the chosen multisource datasets and baseline models to confirm the effectiveness and reliability. The method used in comparison takes into consideration classification accuracy, computational difficulty, and vulnerability to dataset change. Further studies are done on the involvement of various components in the hybrid structure, which produce data on the strong, weak, and potential scaling of the proposed plant disease detection process.

1.6 Scope and Challenges

The research contributes to the computerized identification of plant diseases by addressing the major limitations of analytical and operational capabilities of image-based agriculture diagnosis. The proposed study will create a precise and computationally beneficial deep learning model, which can identify the diseases of plants using leaf images in a proper manner despite the lack of knowledge and resource accessibility. This research aims at enhancing diagnostic validity in situations where there is a lack of labelled data and capacity to compute through the application of transfer learning, hybrid structures, and high-order classification algorithms.

In this section, the reason for the research is explained, and the main challenges of finding the right, effective, and flexible ways of identifying the disease with the help of an image are underlined, along with the main challenges in creating efficient, practical, and effective methods of identifying diseases in real-world situations, i.e., in farming.

1.6.1 Research Scope

The primary goal of the research is to confirm the effectiveness of machine learning-based classification and extraction of feature algorithms to identify plant diseases using information on leaf images. The research concentration is established in a number of factors:
Only RGB and secondary leaf photos are used in the study as the primary medium of data, since the emphasis is on visual assessment of complaints to identify a disease. This method allows developing a light and flexible architecture that can be integrated into the real agricultural surveillance systems, even though additional sensors or environmental knowledge are required.

The study provides numerous pretrained convolutional neural network architectures in order to determine the proper baselines, including ResNet50, VGG16, MobileNetV3, DenseNet121, InceptionV3, and EfficientNet-B0. These models are an embodiment of a few design philosophies as regards depth, parameter efficiency, and receptive field properties, which can be systematically compared under equivalent training conditions.

Along with single-architecture models, the paper introduces the CNN-Transformer architecture, comprising the small MobileNet, Tiny ViT, and EfficientNetB0. The approach attempts to describe local texture-level features and the global contextual features and it defeats the drawbacks of all-convolutional models to describe complex disease patterns.

Transfer of learning is among the main tools of analysis that are applied to overcome data gaps and accelerate convergence. The field is also looking at fine-tuning techniques to achieve models that are formally trained to plant disease settings and maintain generalizability to diverse disease classes. The field also includes the examination of integrated feature techniques, such as the Kolmogorov-Arnold Network (KAN), as an alternative classification methodology. Performance assessment is conducted on the basis of standard classification measures and computation effectiveness measures, as well as an emphasis on flexibility in the limited resource farming cases.

The extent of disease detection, spatial development analysis, field work application, and multiple-modal identification situations are not discussed under this research paper because they are considered to be outside the scope of the research.

1.6.2 Research Challenges

Taking into account current developments in the field of neural network detection of plants, numerous challenges still exist to prevent the creation of reliable and scalable systems in the field of diagnostics. The major challenges in this area include the inadequacy and discrepancy of marked plant disease data, which restrict guided learning and increase the risk of over-fitting, particularly across the rare types of disease classes. The other major issue is that there is a significant visual homology of various plant diseases, especially at the early stages of infection. This inconsistency is brought about by overlapping color, texture variations, and sound in the surroundings, which requires accurate recognition to be difficult, even in the case of deep neural networks.

Another limitation is computational complexity since the large memory and computational power often demand deep learning models with high capacity. This restricts their use in low-cost or edge-based agricultural systems where precise architectural design is necessary to balance between precision and efficiency.

The emergence of hybrid CNN-Transformer systems brings into existence novel questions of successful feature fusion. Also, it is challenging to incorporate complementary information of convolutional and attention-based models without adding redundancy or excessive parameter enlargement. Moreover, the application of Kolmogorov-Arnold Networks (KAN) as classifiers is a methodological issue since they are new to the pipelines of deep learning. Nonlinear decision modeling requires proper tuning, stability testing, and performance verification.

Finally, one of the big challenges is the trade-off optimization between performance, in terms of classification, and computing economy. To ensure that the proposed structure is relevant to be implemented in resource-constrained practical farming scenarios, accuracy should be optimally balanced with the complexity of the models.

Chapter 2

Literature Review

2.1 Preliminaries

Plant diseases are also a major threat to agricultural productivity and world food security especially on staple crops like rice. Disease diagnosis in plants is very important to minimize losses in yields, preserve quality and to promote sustainability in agricultural activities. The general process of disease diagnosis based on traditional disease diagnosis methods is manual examination of the experts, time consuming, subjective and not easy to scale in big farming regions [18], [20]. Due to this, automated image-based disease detection systems have come forward as a viable and scalable alternative.

The recent developments of computer vision and the machine learning field have turned the diagnostic process of plant diseases into a data-driven, automated process. Deep learning frameworks can discern discriminative features directly out of raw photographs, which obviates the use of handcrafted features mover and attains better results than antique image processing and machine learning algorithms [11], [22], [23]. In this part, the background knowledge that will underpin this study will be provided, which are computer vision-based plant disease detection, convolutional neural networks, Vision Transformer networks, transfer learning techniques, and hybrid CNN-Transformer systems.

2.1.1 Plant Disease Detection using Computer Vision

Computer vision methods make it possible to analyse and interpret visual data acquired through plant images with the use of automated methods. These methods are normally deployed in agriculture to identify symptoms of diseases like discoloration, lesions, necrotic spots, and texture anomalies. Old methods used classical image processing and hand-designed features and used traditional classifiers, including Support Vector Machines (SVM) and k-nearest neighbors (k-NN) [20]. As much as these methods performed averagely, their accuracy was very sensitive to the lighting conditions, the complexity of the background and the design of features.Deep learning has tremendously enhanced the resilience and accuracy of plant disease detection systems. Convolutional neural networks (CNNs) are able to learn hierarchical features of images in an automatic way, which allows them to identify minor symptoms of a disease that traditionally cannot be detected. Mohanty et al. [11] showed that deep learning models that are trained on datasets of large leaf images can be highly accurate when it comes to classification under controlled conditions. The efficacy of deep learning in different crops and types of diseases was also confirmed by further research, such as rice [6], [9], [16]. These results make deep learning based on computer vision a trusted source of automated plant disease detection.

2.1.2 Convolutional Neural Networks (CNN) for Image Classification

Convolutional Neural Networks are some of the most popular deep learning constructions in terms of image classification tasks. The CNNs are built upon layers of stacked convolutional, pooling, and fully connected layers, which extractively identifyively more abstract features out of the input images. The low level features, including edges and color gradients, are represented in the early layers while deeper layers interpret the higher level semantic features that involve disease specific textures and gradients.

VGG, ResNet, DenseNet, MobileNet, and EfficientNet are some of the examples of CNN architectures that have been used to detect plant diseases. EfficientNet proposes the compound scaling to trade-off depth, width, and resolution to have good performance at a lower computing expense [1]. MobileNet uses depthwise separable convolutions, which should be used in mobile and embedded agricultural applications [5], [13]. It has been demonstrated that CNN based models with the aid of preprocessing and data augmentation are capable of high accuracy in classification tasks involving rice disease [6], [9], [16]. But CNNs are local Spatial first and might not be able to reflect long range contextual dependencies on the full picture.

2.1.3 Vision Transformer (ViT) Architecture

Vision Transformers are a departure of the convolution-based processing that adapts transformer-based architectures originally created to operate in natural language processing. ViT models also separate images into fixed-size patches, encode them into tokens and handle them with multi-head self-attention mechanisms, which allow the modeling of global contextual relationships [24].Transformers can use long-range dependencies that CNNs struggle to capture because of this design.

Transformer-based architectures have already demonstrated promise in the classification of images, their segmentation, and estimation of the severity of diseases. Vision Transformer has also been applied to rice leaf disease and multitask disease-severity prediction in an agricultural setting with success [2]. Variants of hierarchical and lightweight transformers also utilize less computational resources without impairing their high representational power [12]. These attributes render transformers an important complement of CNN-based feature extractions.

2.1.4 Transfer Learning Framework

A common deep learning technique is transfer learning, which involves fine-tuning models trained on large-scale datasets, such as ImageNet, to domain tasks.This method saves time in training, also it is more convergent and better in performance when there is a lack of labeled data.

Transfer learning has always shown improvements in the performance of several CNN and transformer architectures in the context of plant diseases detection [6], [10], [14]. The use of pretrained models that are fine-tuned enables generic features of visuals to be used but trained to fit disease-specific patterns. Progressive freezing of layers and massive data enhancement also enhance stability and generalization further [14]. In this study, we use transfer learning on all the backbone networks to use the already learned information and enhance their strength on rice leaf disease data.

2.1.5 Hybrid CNN–Transformer Model

Hybrid CNN-Transformer models are developed on the basis of the complementary strengths of transformer and convolutional-based models. CNN parts aim at extracting fine-grained local features (e.g., texture and lesion boundaries) whereas transformer parts learn global contextual relationships with the help of self-attention mechanisms.

Recent literature has demonstrated that hybrid models are better than single CNN or transformer models in the classification of plant diseases. Hybrid CNN-ViT models have been reported to perform better in accuracy, robustness and generalization on various crop datasets [3], [7], [19]. The discriminative capability is also improved by feature fusion methods that combine local and global representations without impacting the computational performance.

Inspired by these results, this study suggests a hybrid architecture based on the combination of Tiny Vision Transformer, EfficientNetB0, and MobileNet and a Kolmogorov-Arnold Network (KAN) classifier. The upcoming design seeks to provide high classification and feature representation, and efficient deployment needs to the automated rice leaf disease.

2.2 Review of Existing Research

2.2.1 CNN-Based Plant Disease Detection Methods

Shoaib et al. [4] provided an extensive literature review of recent changes in deep learning in detecting plant diseases, in which he discussed the drawbacks of the manual process of identification. They concentrate on the machine learning (ML) and deep learning (DL) methods, in particular, Convolutional Neural Networks (CNNs) and Deep Belief Networks (DBNs). Such methods can automatically derive features in images and detect subtle disease symptoms that the traditional methods cannot. They underscore the fact that they are much more effective and precise, especially due to the emergence of transfer learning and ensemble approaches. Amazingly, a residual attention mechanism network was able to achieve 98 percent accuracy on the PlantVillage dataset. The article goes on to discuss how it is possible to use drone technology with R-CNN based instance segmentation which recorded 97.24% accuracy and a 79.31% average IoU on northern leaf spots. Such measures of evaluation as the IoU, Dice Similarity Coefficient (DSC), and the recall curves are discussed. Nonetheless, there are issues such as the need to have big labeled datasets, the quality of the data, the computational capabilities, and failure to distinguish between healthy and unhealthy plants. This paper identifies the significance of public datasets and the continuous nature of the necessity of strong, generalizable models to be used in the real world.

The paper investigates how deep learning methods can be used to detect plant diseases automatically with the help of leaf photos, which would help address the limitations of the existing visual observations and classical machine learning approaches. The manual diagnosis of diseases is time-consuming, subjective, and predisposed to human error, but the previous computational algorithms often fail to generalize properly in real-world settings because of variations in lighting, background structure and the appearance of the disease.Ignoring the shortcomings of the previous computational algorithms, the authors Sennan et al. [3] employed convolutional neural network (CNN) architectures to automatically extract the discriminative characteristics of plant leaf photos. The suggested solution minimized reliance on features generated by hand and maximized resistance to the processing of complex visual patterns related to diverse plant diseases. To address the challenge of little labelled data, data augmentation and transfer learning models were applied to enhance the generalisation of the model. The experimental evidence revealed that deep learning models were effective in a wide range of categories of plant diseases, outperforming the classical machine learning approaches by a value of 97.5% in terms of the area under the curve (AUC) value. Although such positive results were achieved, the paper observed numerous challenges, such as imbalance in datasets, visual similarity between disease symptoms, and the complexity of the computing procedure, which could hinder the real-world application, especially in the agricultural environment with limited resources.Overall, the data demonstrate that the deep learning approach is a promising solution to automated detection of plant diseases. Nevertheless, the paper emphasizes the necessity to develop effective structures, to work with more and bigger information and to develop lightweight models to apply to real farming. The results of this study lead to further studies on effective and hybrid deep learning models, including the use of CNN-Transformers, to enhance the accuracy and deployability of the classification of plant diseases.

Plant diseases are increasing and are impacting on growth among the farmers. It is significant to make it easy through early detection. The precision with which plant diseases are identified using image information based on deep learning models has increased immensely. The work of Chowdhury et al. in which they came up with a modern deep convolutional neural network (CNN) called FL-EfficientNet model, which has less central processing requirements and combines federated learning (FL) with the EfficientNet framework to achieve high classification accuracy [1], is just one contribution. Their system with its strength in controlled settings was applied on a set of leaf images of ten most common plant diseases and attained a high accuracy of 99.72%. FL is an effective substitute to the classical centralized training, and the utilization of FL also guarantees the preservation of privacy over the edge devices and minimizes the transmission expenses. Nevertheless, the experiment was mainly based on clean and well maintained datasets in which the image can be influenced by background clutters, light difference and occlusions to render the model potentially inefficient in the real world setup.Moreover, it overlooked the practical issue and failed to deal with multi-disease diagnosis in agriculture. To eliminate these challenges, the authors came up with a two-stage model that proposed realistic image enhancement in which it first identifies diseased areas and then determines the kind of disease, which may go a long way to enhance its application in the field of application.

2.2.2 Vision Transformer-Based Plant Disease Classification

The MobileNets, an efficient convolutional architecture of neural network to be used in mobile and embedded vision, were proposed by Howard et al. [5]. Their basis is depthwise separable convolutions which break standard convolutions down into depthwise and pointwise (1×1) convolutions resulting in a computation reduction of around 8-9× with negligible loss of accuracy. These authors propose two world-level hyperparameters: the width multiplier (α) that drops the number of the channels in each layer and the resolution multiplier (ρ) that downsizes the input image by a factor of about $\alpha^2\rho^2$, effectively two-dimensional multiplication, to make trade-offs between accuracy, latency and model size. The baseline MobileNet-224 (α=1.0) has 4.2 million parameters and 569 million multiply-adds and 70.6% elucidate-1 exceptionalization on the ImageNet network classification benchmark. It is more accurate and efficient than AlexNet and SqueezeNet, and only 32 times the number of parameters and 27 times less computation costs are needed to achieve the same level of performance as VGG16. Massive studies indicate that smooth accuracy deterioration occurs with a widening of models or an increase in input resolution, which demonstrates the scalability of MobileNet. Moreover, MobileNets are universal in terms of vision tasks, such as object detection, fine-grained classification, face attribute recognition, face embeddings, and large-scale geo-localization, frequently achieving state-of-the-art performance compared to heavier models at at least half the required computational cost and thus MobileNets are best suited in real-time resource-constrained deep learning tasks.

The further development of the classification of plant leaf diseases indicates the increasing popularity of deep learning in order to address the drawbacks of the traditional method of manual inspection and the classical machine learning approaches. Although previous methods used image processing, support vectors machines, random forests, and convolutional neural networks (CNNs) recorded decent performance, the method used in these techniques was constrained by the use of the local receptive fields, which limited the capability of the techniques to learn long-range interactions of intricate patterns of leaf diseases. To overcome these shortcomings, global self-attention mechanisms of transformer-based models have been applied in the computer vision field. Han and Guo [10] proposed a hierarchical vision transformer-based model together with transfer learning to automatically classify the ligneous leaf diseases. The model uses a two channel swin transformer structure to derive complementary features of original leaf images as well as respective edge images produced by the Sobel operator. Through hierarchical feature extraction, shifted-window self-attention and ImageNet pre-trained weights, the given approach allows stimulating classification performance with least computational complexity. Experimental tests on a published dataset consisting of 22 classes of ligneous leaf diseases prove that the transformer-based model is better than state-of-the-art CNN and traditional vision transformer models in accuracy, precision, recall, and F1-score. On the whole, this paper confirms the usefulness of hierarchical vision transformers to the classification of plant diseases and shows that they can be used in precision farming, especially, in those cases, when global contextual information is important.

Latif et al.[6] introduced a deep learning-based system to detect rice plant diseases by a modified VGG19 transfer learning CNN, which is of great importance as it will allow detecting disease in the early stages before it can lead to a loss of 20-40% of crops every year. The paper is devoted to the classification of six categories of rice leaf conditions as healthy, narrow brown spot, leaf scald, brown spot, and bacterial leaf blight in terms of a publicly available rice leaf image dataset augmented by image preprocessing and intensive data augmentation (rotation, translation, and scaling) to curb overfitting due to small volumes of data. The authors tested various popular CNN models such as GoogleNet, VGG16, VGG19, DenseNet201 and AlexNet and showed a significant performance increase in all models using transfer learning. The most successful of them was the fine-tuned VGG19 model, which reached its highest classification accuracy of 96.08% in the case of the non-normalized augmented dataset with a corresponding precision of 0.9620, recall of 0.9617, specificity of 0.9921, and F1-score of 0.9616, which was superior to existing methods of detecting rice disease reported in the literature as most of them used smaller datasets and fewer disease classes. Analysis of experimental data also supported the training behavior at the stable with less overfitting since the accuracy and loss curves in training and validation align. The article states that the suggested VGG19-based transfer learning model offers a robust, precise, and scalable solution to the problem of rice leaf disease diagnosis and is a good perspective of its application in practice when combined with IoT-based and drone-based surveillance, which is why it could become the leading model in future studies on the topic of agricultural disease detection.

The diseases associated with rice are highly precise within the system of deep learning-based hybrid ensemble architecture introduced in Islam and Richhariya [8] related to the significance of rice in world food security and the vast losses to the crop caused by plant diseases attacks. The paper operates with a massive, publicly available, Kaggle dataset of 11,790 annotated images in nine classes of rice leaf disease (brown spot, leaf blast, hispa, neck blast, and others) to take a series of systematic steps taken to process images (resizing to 256×256, decoding, normalization, noise elimination, and set separation 70:10:20) in order to make the model more robust and use it as the foundation of the research (11,790 annotated images). Several transfer learning models ResNet152V2, DenseNet121, InceptionResNetV2, and MobileNetV2 pre-trained on ImageNet were trained as feature extractors, with DenseNet121 and ResNet152V2 selected for constructing a hybrid ensemble architecture due to both dense and residual feature learning qualities that were complementary to each other and trained on ImageNet as feature extractors. The two networks were similar and were trained with frozen weights and without the top layers, extracted features were batch-normalized, flattened, concatenated and then a nine-class softmax output added. The model was trained on Adam optimizer and categorical cross-entropy loss, which were evaluated with the assistance of accuracy, precision, recall, and F1-score. The results of the experiment have revealed that the developed model of the hybrid ensemble has demonstrated high and consistent results with an accuracy of 98 and a precision, recall and F1-score, which are by far better than the single-model-based model approaches and the methods previously used that are based on CNN with reported accuracies of between 95 percent with smaller dataset sizes. The comparative analysis and the confusion help to demonstrate the high generalization ability of the model and the reliability of the model concerning the classes which proves the efficiency of the ensemble learning and transfer learning in the diagnosis of the rice leaf disease. In the study, the proposed solution is a scalable, accurate, and robust solution to automated detection of rice disease, which can be applied in a substantial way to the real world to detect the diseases in agricultural fields and extended to estimating the disease severity, disease progression, and reinforcement learning-based crop health monitoring.

Panchal et al. [9] developed a plant disease detection system based on the images in an agricultural environment using deep learning to make precise and automated diagnoses. Their approach involves the use of transfer learning and custom CNNs applied to a pretrained model ResNet50, VGG16 and Inception-v3 on the plantvillage data, consisting of approximately 50, 000 RGB images of 14 different crops and 38 disease categories. To enhance the generalization of the model, preprocessing was done using resizing, normalization, and data augmentation. The validation accuracy of VGG16 is 93.5% and it is better than other models due to its powerful feature extraction features. Comparatively, a 10 layers custom CNN achieved only 81.4% which demonstrates the power of transfer learning in small-data settings. ResNet50 was not effective in this plant classification task, whereas it is generally more effective with large datasets. According to the stage-wise analysis, the number of leaves that were identified as healthy was the most ideal whereas accuracy was lower in the intermediate disease stages, frequently due to similar symptoms. The paper acknowledges its limitations such as threats of overfitting, unequal imaging conditions and dependency on datasets. With this, it has a flexible, low complexity solution that can be deployed on mobile. The authors will incorporate decision-support tools and environmental sensor data in future, enhancing its realistic usefulness to the farmers in resource constrained real world scenarios.

2.2.3 Hybrid CNN–ViT Approaches in Agriculture

The recent studies have also examined the methods of deep learning to enhance automatic detection of plant diseases, particularly when it comes to economically important crops such as rice. The existing disease diagnostic processes that are based on manual visual examination are time-consuming, subjective, and inaccessible to the smallholder farmers. Despite some promising results of the previous machine learning and convolutional neural network (CNN)-based approaches, they are often crippled by overfitting, high computing cost and incapability to sufficiently represent the long-range spatial correlations in the leaf pictures. To overcome these limitations, Roy and Kukreja [2] introduced a ViT-based model that can be used to both identify rice leaf diseases and determine the severity of the disease at the same time. The given approach employed a learning strategy of multitask which was based on the concept of self-attention to model global contextual relations in leaf images. The analysis presented a custom dataset of 3,345 annotated rice leaf images of ten types of diseases, and three levels of severity: mild, moderate, and severe, to enable better representation of rich and long-range visual features Unlike conventional CNN models, the ViT model split input image into fixed-size patches and performed multi-head self-attention and positional embeddings, which enables more severe and complex visual patterns to be represented. A common backbone architecture was created to simultaneously learn the kind and severity of disease and reduce computing overhead and enhance feature reuse. The model was also trained with cross-entropy loss and Adam optimizer and data augmentation strategies were employed to enhance generalization. Performance was evaluated using precision, recall, the F1 score and the area under the receiver operating characteristic curve (AUC), the experimental results demonstrated moderately good accuracy in disease classification (weighted F1-score of approximately 54%) and significantly better results in severity estimation (weighted F1-score of almost 78% and an AUC value of 0.86). Some classes of diseases like Yellow Molten were more easily classified but physically similar diseases like Rice Blast were hard to separate. It was also reported that moderate and severe level of severity were oftentimes not correctly classified, which is why the authors emphasized that the feature discrimination and variety of the dataset could be improved in the future to support transformer-based designs in analysing pictures of agricultural objects.Overall, the research contributes to the growing literature that backs the idea that transformer-based designs are good at analysing pictures of agricultural objects. It shows that Vision Transformers can be as viable alternative to CNNs in detecting joint illnesses and assessing their severity due to its scaling and efficiency. However, the researchers noted that there exist considerable limitations, including limited data size, the presence of skewed classes, and low performance in cases of visual overlap between diseases. The results provide a stimulus to explore the hybrid CNN-Transformer architecture, augment the data more, and enhance explainable artificial intelligence methodology to enhance the trustworthiness and intelligibility of the plant disease diagnosis systems.

Dhakyanaikabde and J [7] is a hybrid deep learning model, that brings together Convolutional Neural Networks (CNN) and Vision Transformers (ViT) to detect the disease in plant leaf accurately and scale-free, overcoming the limitations of the traditional machine classifier and single deep learning models. The presented CNN-ViT network integrates CNN-based local spatial features receptiveness with the global self-attention system of ViT using a channel-aligning feature fusion plan, which can allow the model to adjust both the fine-grained lesion textures and long-range contextual facts. The structure utilizes the image preprocessing methods such as resizing, normalization and data augmentation and then CNN feature extraction and ViT patch embedding and attention learning, where the final multiclass classification is done with a softmax layer. The hybrid model at the PlantVillage banana leaf disease subset with a batch size of 32, a learning rate of 0.001, and 50 training epochs resulted in a mean classification accuracy of 94% ± 0.3, which is better than standalone CNN (86.2%) and ViT (88.1) models ,with improvements of 8.4% in accuracy over traditional machine learning methods and gains of 7.1% precision and 6.5% recall compared to ViT.. Ablation studies and confusion matrix established robustness in classes and complementary nature of CNN and ViT elements whereas computational profiling indicated the inference latency was reduced by 23% and memory usage decreased by 18% than transformer alone models, which enabled edge-level deployment. These findings confirm the suitability of the suggested hybrid CNN-ViT framework to real-time, scalable, and precise leaf disease detection and indicate its possible application to precision agriculture, and possibly other crops, real-field setups, and resource-limited agricultural systems.

Dai et al.[12] introduced a research paper AISOA-SSformer, an enhanced Transformer-based semantic segmentation model of rice leaf pest and disease detection, which can be utilized to overcome the major challenges of the precision agriculture process, such as irregular disease spots, blurred lesion boundaries, and cluttered natural backgrounds that constrain the traditional manual inspection and current deep learning models The proposed model is based on the Segformer-b0 baseline and adds three crucial innovations: (1) a Sparse Global-Update Perceptron (SGUP), replacing the traditional MLP, the combination of stochastic exponential and weighted moving averages in a sliding window to stabilize the parameter changes and achieve better results in small lesion margins; (2) a Salient Feature Attention Mechanism (SFAM), made up of a Spatial Reconstruction Module (SRM) and a Channel Reconstruction Module (CRM)) to suppress background noise and emphasize disease-relevant spatial and channel features; and (3) an Annealing-Integrated Sparrow Optimization Algorithm (AISOA) that extends AdamW by incorporating stochastic sparrow search and simulated annealing to escape local optima and improve segmentation of fuzzy disease boundaries.The model was trained on a small but carefully annotated homemade dataset of 2,005 rice leaf images (1,005 Tungro and 1,000 Brown Spot) resized to 256×256 and augmented with photometric, geometric and scale transformation and trained on an NVIDIA RTX A5000 GPU using PyTorch. It has been shown through experimental results that AISOA-SSformer is able to achieve 83.1% MIoU, 80.3% Dice coefficient, and 76.5% recall using only 14.71 million parameters and 3.28 GFLOPs, and it outperforms models based on CNN, hybrid models (DeepLabV3, MC-UNet), and models based on Transformer (Segformer, Swin Transformer, SETR) in segmentation accuracy at the same level of computational efficiency. In in-depth studies regarding ablation, it has been demonstrated that SFAM offers greatest performance increase, and SGUP and AISOA enhance stability and edge accuracy, and generalisation trials on unknown rice diseases (bacterial blight and blast) exhibit better adaptability, although with small training sample sizes. The suggested framework was also implemented in a real life application system of PyQt-based rice disease segmentation, which proved practical use. In general, AISOA-SSformer demonstrates high accuracy, robustness and efficiency in the process of rice leaf disease segmentation, and therefore it is a good candidate for intelligent agricultural disease management and implementation of edge-devices in the future.

Wang et al. [15] worked on presence challenges to comprehend the intensity based on high inter-class harmony and low inter-class varieties and discussing it, the work suggests CNN based models out of the black plantvillage Dataset, apple black root image which has been categorized into four stages of intensity. Based on a shallow network, the trained and deep models (VGG16, VGG19, Inception, ResNet50) were tested using trans-length. There have been many to four models assessed in an outward manner. Consequently, 79.3% have become precise, however, deep networks have also been fitted because of small data. The good performance of transfer learning has reached 90.4 percent accuracy in VGG16 model which has surpassed some models. Recent Figure-based evaluation method offers the more compatible and accurate results. Even such tools as CLS Rater and applications as Leaf Doctor enhanced the precision. The background removal and the manual type effects are, however, reliant on the semi-automatic techniques. These classical methods of processing images have drawbacks because of the size of the injured areas, color, and the shape variations. Deep learning process offers an automatic resolution which directly gives the characteristics of the properties and threshold based division, rather than pick up on the image information. This approach was demonstrated to be successful in the task of disease detection, identification of plant species and plant diseases including PlantCLEF as seen in the contest. This paper demonstrates that learning, particularly using VGG16 pre-trained models such as VGG16 is promising. The methodological transfer learning that is correctly and automatically adopted on such models is practical to the severity of the plant disease.

2.2.4 Rice Leaf Disease Detection Studies

Kiratiratanapruk et al. [13] showed that deep learning-based object detection could effectively diagnose rice disease, but it is still difficult to apply to the field because of its variability, uneven shooting distances, and considerable differences in the size of objects in a field. The previous CNN-based methods mainly depended on manipulated datasets or structure like feature pyramids and multi-scale attention that enhanced the accuracy of detection with the network at the cost of the computation and complexity. The recent study to counter the variability in object size without modifying network designs proposed adaptive image tiling approaches based on the estimated leaf sizes. The model employed in this way to estimate the rice leaf width is a lightweight CNN regression model (ResNet18), and was applied to allow splitting an image into subimages of equal scale. An experimental assessment of real field pictures of eight rice diseases showed that tiling guided by leaf width was a highly effective way of enhancing detection of various models, including YOLOv4, YOLOv8, and transformer-based detectors, including DINO and Co-DINO. This increase in mean average precision (up to 3.58% in the case of real-time YOLO models and over 94% in the case of transformer-based models) was achieved and was compatible with existing architectures. Transformer models were more accurate, but their computational cost meant that real-time use could not be done, and these Yolo-based models provided a more reasonable tradeoff between speed and accuracy. In general, the literature suggests that adaptive preprocessing methods and, in particular, the object-size-dependent image tiling is a feasible and scalable solution to the problem of enhancing the detection of rice disease in real-world agricultural environments.

According to a study by Vasantha et al. [14], rice leaf disease detection indicates the increased relevance of the approaches based on machine learning (ML) and deep learning (DL) to early, precise, and automatic diagnosis of crop diseases. Classical ML algorithms such as k-NN, SVM, decision tree, and minimum distance classifier with traditional image processing have demonstrated moderate to high accuracy (0.81 -0.97) with the identification of the common rice diseases, including brown spot, leaf blast, sheath blight, and bacterial leaf blight. Recent developments indicate that the deep learning methods, specifically convolutional neural networks (CNNs) and its variations, including CNN-SVM hybrid, transfer learning, Faster R-CNN, and deep feature-based classifiers, reliably surpass the traditional ones, with performance reaching 0.98-1.00 in various disease subsets. It has also been reported that lightweight architectures, and ensemble or fusion-based CNN models can enhance robustness and high performance. Despite the fact that classical ML models are still competitive in the case of small disease categories and a small number of datasets, CNN-based models and transfer learning can be used to offer better generalization and scalability to real-world applications. All in all, it can be concluded that deep learning-based image-based disease detection is a viable and efficient approach to monitoring the health of rice crops, especially in arid and extensive agriculture systems, where resources are limited.

2.2.5 Lightweight and Efficient Deep Learning Models

The proposed lightweight deep convolutional neural network (DCNN) to train on recognizing the key rice pests and diseases is presented by Zhang and Yu [17] based on the MobileNet-V2 backbone that includes the three main purposeful modules, i.e., a super-resolution (SR) front-end that improves a low-quality field image, an adaptive spatial-channel attention module (ASAM) that focuses on the most critical lesion and pest features, and a cross-level feature fusion module (CLFFM) that integrates the low-level features, Subset-wise analysis also shows the robustness across all the cases (public and field images), higher performance than MobileNet-V2 ($p < 0.01$) and device-level profiling indicates the near real-time inference with Jetson (Nano, INT8) and on Raspberry (4B, FP32) devices. The confusion matrix analysis ensures that there is a high inter-class differentiation especially when dealing with more complicated types of pests such as Brown Planthopper (95.3%) and Leaf roller (93.6%) and fewer misclassifications are made when the lesions are still young or when there is interference with the environment. The sources of SR, ASAM, and CLFFM, as well as the benefits of ASAM compared to other variants of attention, are confirmed by the work on ablation, which demonstrates a complementary influence on accuracy, which does not require considerable overhead. Although these strengths are present, it has limitations such as there may be a domain shift between geographies, cultivars, and acquisition devices, inter-class feature ambiguity at similar disease stages, limited open-set recognition, there may be SR/attention failure under harsh light or motion, it is hardware-dependent, and imbalance of labels. Future directions will be based on open-set recognition and anomaly detection, domain adaptation of cross-region / cultivar implementation, and the initial symptom identification to make agricultural robotic systems more robust, safe, or more general.

Rice diseases are very serious because they affect agricultural output and food security in the world hence, it is important that they are detected early. The image-processing techniques that are traditionally based on the RGB values are not always reliable when used to detect and classify rice diseases. Convolutional Neural Networks (CNNs) and machine learning are an optimistic opportunity to address this issue. The paper of Hasan et al. [16] demonstrates that the models of deep learning are often associated with high-performance hardware that may consist of costly GPUs, which adds to the cost of operating them. This paper presents a light CNN model that has moderate- high accuracy in the classification of rice leaf diseases with minimum computation requirements. Segmentation and K-means clustering are applied to do pre-processing of rice leaf images to isolate the affected areas by removing background and green areas. The model is trained on an augmented set of 2,700 images and tested with 1,200 images and a testing accuracy of 97.9 percent on rice brown spot, bacterial blight and leaf smut. Ultimately, the trained model is implemented in a mobile application, which proves the applicability in the real world. The efficacy of the proposed approach regarding accuracy, efficiency, and practicability is proven by the comparison with the existing works.

Ali et al. [19] introduce a small deep convolutional neural network that can be used to identify rice pests and diseases, with MobileNet-V2 as its backbone, a super-resolution (SR) front-end to boost the quality of low-quality field images, an adaptive spatial-channel attention (ASAM) module, and a cross-level feature fusion module (CLFFM). The proposed model is evaluated on a combined dataset of RiceLeaf, PlantVillage and field collected images and yields 94.3% and 0.938 accuracy and F1-score, respectively, exceeding mainstream architectures like ResNet-50 and EfficientNet-B0, and maintains a small model size of 9.7MB and low inference latency of 11.2ms (input 224×224, batch size = 1). Subset-wise analysis demonstrates that the robustness is similar in both public and field images and statistically significant gains over MobileNet-V2 ($p < 0.01$). Device-level micro-benchmarks validate close to real time performance for Jetson Nano ($\sim$33 ms, INT8) and interactive latency for Raspberry Pi 4B ($\sim$118 ms, FP32), which can be deployed in edge and mobile applications. The evaluation of the confusion matrices demonstrates a high inter-class separation, particularly in complex classes of pests, including Brown Planthopper (95.3%) and Leaf Roller (93.6%), although misclassifications are mostly viewed in young lesion stages or through interference of the environment. Ablation experiments confirm the efficacy of SR, ASAM and CLFFM and point out the superiority of ASAM compared to other attention mechanisms, where complementary gains are attained without much overhead. These are characterized by domain shift between regions and cultivars, inter-class feature ambiguity in similar disease-stages, inability to perform open-set recognition, possible failure at low-light or motion conditions in SR/attention, region-dependent performance, and label bias. Future research will focus on open set recognition and on anomaly detection, domain adaptation for cross-region and cross-cultivars deployment, and symptom detection in early stages to be able to increase robustness, generalization, and practical applicability in the agricultural setting.

2.3 Summary of Key Findings

The review of the research on detecting plant diseases based on image data shows the need for quick and steady improvement interconnected with the development of deep learning networks, transfer learning, and optimal model design. The initial methods were based on the traditional image processing and the classical methods of machine learning (SVM, KNN, ANN) which used handcrafted color and texture cues. Although these techniques were moderately successful, they were also not easily scalable, strong, and capable of adapting to more complicated field environments.

Since the development of convolutional neural networks (CNNs), the accuracy of detecting plant diseases has significantly increased because of the automaticity of the process of feature extraction and the hierarchical representation learning. VGG, ResNet, DenseNet, and EfficientNet among others exhibited good performance on various crops and disease types, with most achieving much more than 95 percent accuracy on typical crop and disease publicity like PlantVillage. Transfer learning also improved the results in low-data conditions, decreasing training time and computational cost and increasing generalization.

Recent studies have been conducted to extend beyond the standard CNNs and move towards more efficient, scalable and context aware architectures. FL-EfficientNet represents a federated learning-based model that considers privacy issues in data and communication because it allows edge-based training. Lightweight networks such as MobileNet and MobileNet-V2 support mobile and embedded systems, and therefore, real-time disease diagnosis is possible on farmers operating under resource-constrained settings.

The other significant development is transformer-based models, especially the ViT and CNN-ViT hybrid models. These models are able to use self-attention mechanisms to model long-range spatial dependencies and global contextual information which CNNs may have significant difficulty modeling. Multitask learning methods that integrate disease classification and severity estimation have added usefulness to practical applicability particularly to crops such as rice whose disease progression has a direct impact of crop yield.

Pipelines and attention mechanisms which are assisted by segmentation have also become prominent. Techniques that involve the use of U-Net based segmentation, attention modules (CBAM, ASAM) and semantic segmentation transformers (e.g. SegFormer based frameworks) are more robust in isolating diseased regions and removing noise in the background. These plans are very useful in enhancing performance in adverse environments like disorganized backgrounds, irregular shapes of lesions, and in the light changes.

All in all, the literature suggests that there is a marked shift in the types of models that once emphasized accuracy and computation-intensive models to the ones that are efficient and interpretable and can be used in real-world farming environments.

2.3.1 Major Findings and Trends

Several key trends emerge from the reviewed studies:

Replacement of handcrafted characteristics by deep features learning: The historical ML tools with manual feature extractions are being substituted by CNN-based architectures that acquire discriminative features directly on images, which leads to high accuracy and robustness.

Adoption of transfer learning and data augmentation: Most high performing models make use of pretrained CNN backbones (e.g. VGG19, ResNet50, DenseNet121) and aggressive augmentation in order to alleviate limited and imbalanced datasets.

Move to efficient and light weight models: With architectures like MobileNet, EfficientNet, and minor customized CNN, which reduces the number of parameters and lower the inference latency and preserves the accuracy for mobile and edge applications.

Combination of attention and segmentation processes: Attention modules and segmentation-based preprocessing help to enhance focus on the disease-relevant regions and reduce the interference of the background and enhance class separability.

Emergency of transformer & hybrid cnn-vi models : Vision Transformers and CNN-ViT hybrids discover global associations and contextual associations, and they are better than standalone CNNs in challenging disease patterns and severity prediction problems.

Concentrate on the applicability in the real world: computational efficiency, and execution of the methods on smartphones, drones, and edge devices, considering the reality of agricultural constraints.

Although these improvements have been made, some of the issues persist such as bias in datasets to controlled settings, inability to generalize to real-field conditions, visual similarity of diseases, and open-set recognition.

Table 2.1: Comprehensive Review of Plant Disease Detection Using Image Data

Study / Model	Dataset	Key Contribution	Limitations
Chowdhury et al. [1]	Leaf images (10 diseases)	Federated learning with EfficientNet for privacy-preserving training; achieved 99.72 percent accuracy on controlled datasets	Trained on clean datasets; limited robustness to real-field conditions; multi-disease coexistence not addressed.
Roy and Kukreja [2]	Rice leaf images (3,345)	Vision Transformer with multitask learning for disease classification and severity estimation	Small dataset; difficulty distinguishing visually similar diseases
Sennan et al. [3]	Multi-crop leaf images	CNN-based automatic feature extraction; data augmentation and transfer learning improve generalization; AUC about 97.5 percent	Dataset imbalance; high computational complexity
Shoaib et al. [4]	PlantVillage and UAV datasets	Comprehensive review of CNN and attention-based models; residual attention networks achieved 98 percent accuracy	Requires large labeled datasets; high computational cost; limited distinction between healthy and unhealthy plants
Howard et al. [5]	ImageNet	MobileNet architecture using depthwise separable convolutions for mobile vision	Lower accuracy than heavy CNN architectures
Latif et al. [6]	Rice leaf images	Transfer learning with VGG19 improved disease detection accuracy to 96.08 percent	Limited real-field validation
Dhakyanaikabde and J [7]	PlantVillage (Banana)	Hybrid CNN-ViT framework combining local and global features	Requires careful feature alignment and fusion

Study / Model	Dataset	Key Contribution	Limitations
Islam and Richhariya [8]	Rice Kaggle dataset (11,790 images)	Hybrid ensemble of DenseNet and ResNet achieving 98 percent accuracy	Increased model complexity
Panchal et al. [9]	PlantVillage	Transfer learning based multi-crop disease classification	Risk of overfitting and dataset bias
Han and Guo [10]	Ligneous leaf disease dataset	Hierarchical Swin Transformer with edge-aware learning	Requires large datasets for stable generalization
Mohanty et al. [11]	PlantVillage (54,306 images)	Demonstrated feasibility of smartphone-based disease detection	Limited real-world generalization
Dai et al. [12]	Rice leaf images (2,005)	Transformer-based semantic segmentation with attention mechanisms	Moderate dataset size
Lightweight CNN [16]	Rice leaf images	Efficient CNN deployed on mobile devices with 97.9 percent accuracy	Limited disease categories
MobileNetV2 with Attention [17, 19]	Mixed public and field datasets	Lightweight attention-enhanced CNN for real-time deployment	Domain shift and label imbalance

2.3.2 Implications for the Future

The further evolution of the efficient deep learning-based models of the detection of the plant diseases is of great importance to the fields of precision agriculture, food security and sustainable farming. Early and correct diagnosis of the disease can lead to the timely prevention of the disease, minimize crop loss, and avoid excessive use of pesticides.

Future research directions include:

- Training powerful models over a wide range of real-field data.
- The architecture of mobile deployment of lightweight and accurate architectures.

- New hybrid CNN-Transformer systems using superior contextual knowledge.
- Multi-disease, multi-stage, multi-severe multi-disease models.

Addressing these questions, systems of disease detection that are scalable, reliable, and acceptable to farmers, and can be operated in a real-world agricultural setting will be obtained.

2.3.3 Research Gap

Although the automated plant disease detection using CNNs, Vision Transformers, and hybrid CNN-ViT structures have made significant progress, there are a number of key gaps:

Models which have been developed are highly accurate when used on curated datasets like PlantVillage but they fail significantly when exposed to uncontrolled field conditions. This deterioration is due to the change of light, occlusion, clutteriness of the background, and the occurrence of multiple diseases at the same time.

The challenge of visual similarities between diseases, like those between leaf blast and brown spot, has been a major problem to CNN-only and transformer-only architectures in the early stages of infection since such variations in texture and hue can be the only factors to be considered to identify the exact disease.

Transformer-based models and deep CNNs require large amounts of memory and processing power which limits their use on edge devices or mobile agricultural systems that require lightweight and efficient architecture support.

Even though the hybrid CNN-ViT models are successful in learning both the local and global features, their combination with other advanced classifiers like Kolmogorov-Arnold Networks (KAN) is not well studied. The ability of KAN to represent nonlinear and complex decision boundaries in multi-class plant disease detection has not been fully explored.

Although the use of Grad-CAM, LIME, and attention mechanisms has existed in isolated studies, not many have systematically investigated and confirmed the assertion that models are focusing on disease-relevant areas and not arbitrary background objects, which is essential in forming confidence in model use in the real world.

There is also an evident necessity to have a lightweight and hybrid CNN-ViT model combined with KAN, which can be optimized to achieve high computational efficiency and classify visually similar multi-class rice leaf diseases, and has strong performance in the field along with explainable predictions.

Chapter 3

Research Methodology

3.1 Methodology Overview

In this study, they suggest a unified and effective deep learning model to identify multi-class rice leaf disease using image information. The methodology employs a risk-reducing and systematic working procedure by beginning with controlled data collection and preprocessing, then moving to learning hybrid models and ensuring experimental reliability, reproducibility, and computational efficiency. Architectural analyses and ablation research are performed in order to analyze performance, generalization capability, and computational cost.

The classification of rice leaf disease is challenging because of the great inter-class visual similarity, the presence of similar symptoms between different stages of the disease, and changes in the illumination and the background condition. Single-architecture models can be difficult to both learn fine local lesion detail and long-range contextual patterns on the surface of a leaf. To overcome these drawbacks, this paper uses a hybrid deep learning approach that combines complementary feature extractors, which allow more robust and discriminative representation learning.

To represent local image texture and global contextual features jointly, the network is a combination of transformers and all the constituents of traditional convolutional neural networks (CNNs). The performance benchmarks are established through the assessment of the baseline models of architecture CNN (VGG16, ResNet50, DenseNet121, InceptionV3, MobileNet, and EfficientNetB0).Further, based on these baselines, a hybrid feature-fusion model based on TinyViT-21M, EfficientNetB0, and MobileNetV3-Small is suggested to be more robust. The classification head is a Kolmogorov Arnold Network (KAN) that is effective in transforming complex nonlinear decision boundaries. The training of the model is the result of the two-step approach of transfer learning, which implies stabilizing features and fine-tuning.

The offered system is designed into five modular steps, namely, image acquisition, preprocessing, feature extraction, hybrid deep learning-based classification, and disease prediction. Images of rice leaves are obtained in various publicly available collections to reproduce the variability of appearance of the disease. In preprocessing, images are normalised and resized to 224×224 pixels, and data augmentation is only performed on the training set, to enhance generalization and solve the bias in the classes. The developed images are then fed to several lightweight feature extractors, and the combined representations are then fed to the KAN-based classifier that provides the ultimate disease prediction.

3.1.1 Problem Definition and Research Objectives

Rice is a staple crop of Bangladesh and most other countries. Rice plants are however very vulnerable to many diseases like Leaf Blast, Brown Spot, Bacterial Blight, Narrow Brown Spot, Leaf Scald, Neck Blast, Sheath Blight, Hispa and Tungro. The identification of disease manually is manual and subjective and cannot be used to monitor diseases on a large-scale. The issue that needs to be solved in the given research is automatic recognition and classification of rice leaf diseases on the images into ten classes (nine disease classes and one healthy class). The main objectives are:

- To create a Deep learning-based system that will classify rice leaf disease using automated methods.
- To develop a hybrid architecture that will incorporate TinyViT, EfficientNetB0, and MobileNet. To use transfer learning and fine-tuning to achieve better performance.
- To measure the suggested model with the help of generalized classification measures.
- To consider the efficiency of computations to be deployed.
- To develop an automated deep learning-based system for rice leaf disease classification.

3.1.2 Overall System Workflow

The general sequence of the proposed system functioning is as follows:

- Dataset acquisition
- Data cleaning and preprocessing
- Data augmentation (training set only)

- Feature extraction using TinyViT, EfficientNetB0, and MobileNet
- Feature fusion
- Classification using KAN-based layers
- Performance evaluation

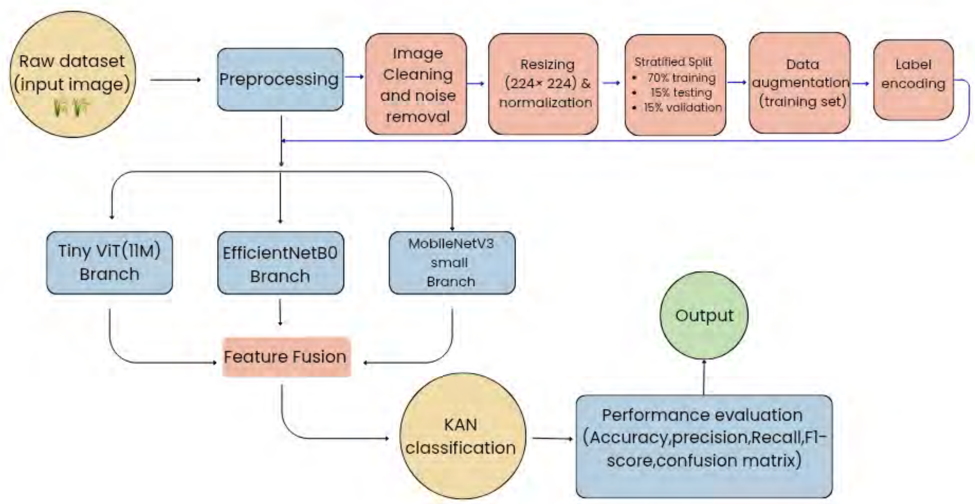

Figure 3.1: Work Flow Diagram

3.2 Dataset Description

3.2.1 Data Source and Dataset Characteristics

The characteristics of data source and data set used are as below: The data required in this study is in the form of rice leaf images that have been gathered in the agricultural repositories and benchmark data that have been put in publicly available repositories. The data set includes the images of both healthy and diseased leaves in different conditions of light, backgrounds, and different leaf postures. The number of images is 23265. All the pictures are classified into ten categories:

- Leaf Blast

- Brown Spot
- Bacterial Blight
- Healthy
- Narrow Brown Spot
- Leaf Scald
- Neck Blast
- Sheath Blight
- Hispa
- Tungro

3.2.2 Class Distribution

The dataset exhibits moderate class imbalance, reflecting real-world disease occurrence patterns. Instead of artificially balancing the dataset, a stratified train–validation–test split was employed to preserve class distribution. Data augmentation and class-aware sampling were applied only during training to mitigate imbalance effects.

3.2.3 Data Splitting Strategy

The dataset is divided using a stratified split as follows:

- 70% training
- 15% for validation
- 15% for testing

This split ensures unbiased evaluation and stable model training.

3.3 Data Collection and Preprocessing

3.3.1 Image Cleaning and Noise Handling

All the collected photographs of plant leaves were carefully observed before training a model so that the quality of the data could be guaranteed, and the experiments could be repeated. Bad files, duplicated samples and poorly labeled photos had been identified and removed out of the dataset. This cleaning process contributed to noisy data cleaning, less dataset bias, and avoiding data leakage between training and evaluation data sets. Only the high-quality and valid photos that had noticeable visual illness characteristics were retained in order to be experimented. This ensured that the learning models concentrated on meaningful disease-related patterns that enhanced the classification performance, resilience and generalization.

3.3.2 Image Resizing and Normalization

The plant leaf image was rescaled to 224×224 pixels in order to fit to the input specifications of pre-trained deep learning models. Instead of scaling to [0,1] pixel values were scaled with model-specific preprocessing techniques that were related to ImageNet-trained weights. In this approach, the distribution of input is balanced to the expectations of the pre-trained models, which have made it possible to extract features more effectively, converge faster, and achieve a higher classification.

3.3.3 Data Augmentation Technique

The training set was only augmented with data in order to enhance the generalization and reduce overfitting. The operations of augmentation were as below:

- Random rotation
- Horizontal flipping
- Width and height shifting
- Scaling and zooming
- Brightness and contrast adjustment

Test and validation sets were not manipulated to provide an unbiased evaluation.

3.3.4 Data Encoding

Disease labels that are in categories were converted to the numerical values with the use of a Cate method that assigns a numerical code to each of the ten disease classes and the one healthy one. This approach allows using more than two classes and makes it possible to use the categorical cross-entropy loss.

The experiment applied three channel RGB tensors to depict all of the images since this approach preserved important color data that were required to distinguish different disease symptoms that encompassed lesions and discoloration and texture alterations. There was consistent label-image matching in all three data sets that involved training and. validation and test sets in order to achieve correct supervision on model training.

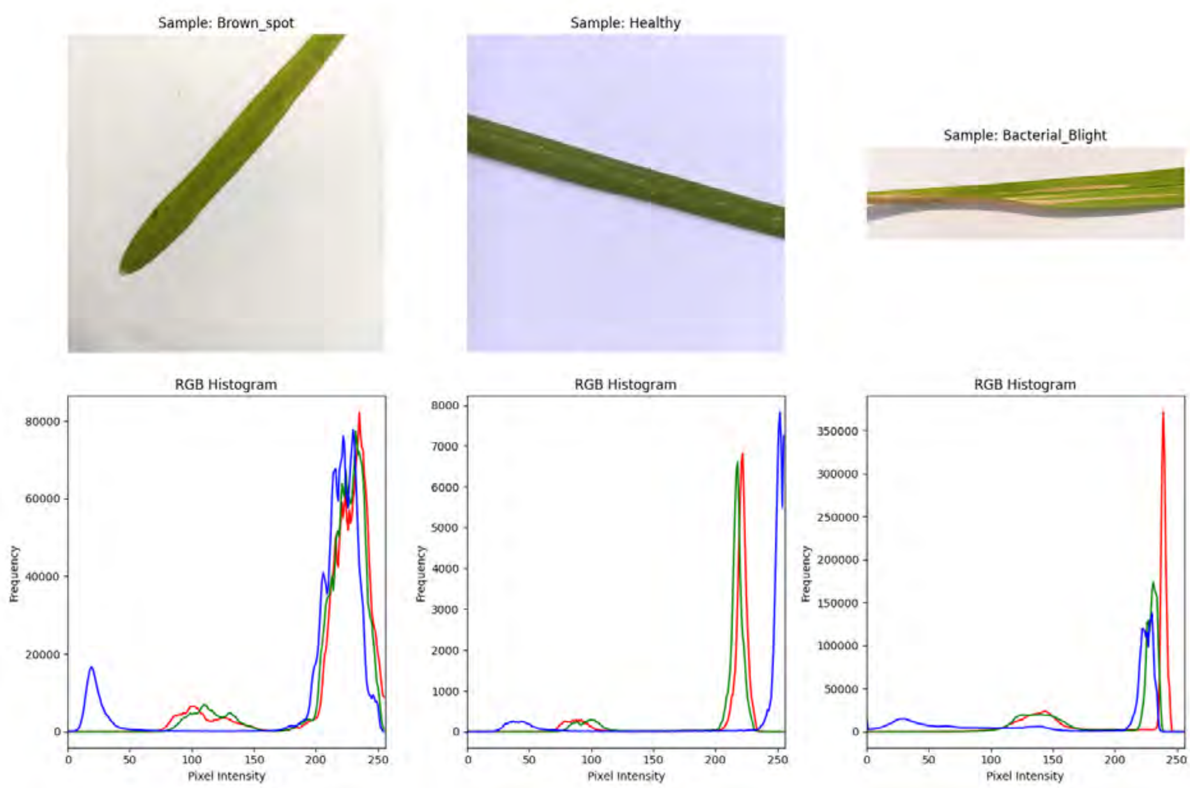

Figure 3.2: CNN sample color histograms

3.3.5 Summary of Preprocessing Pipeline

Summative of Preprocessing Pipeline. The rice leaf images were filtered out to remove rotten, duplicate and mislabeled pictures of the rice leaf. ImageNet-specific preprocessing was applied to the image to resize the image to 224×224 and normalize the image to fit the pre-trained networks. The training set was used to augment the data and increase the level of generalization, but the validation and test sets were not modified. One-hot encoding was used to encode the class labels and all the inputs were also kept in the form of three channel RGB images to retain disease-relevant visual features.

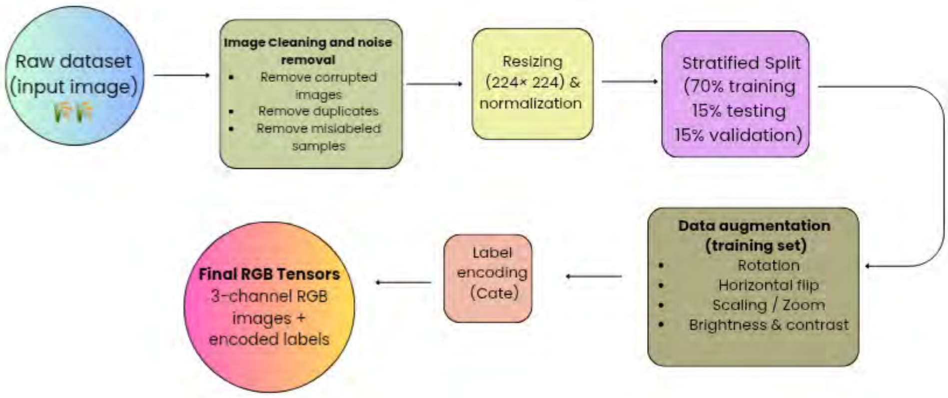

Figure 3.3: Data Preprocessing Pipeline Structure

3.4 Baseline Model Overview

In order to create high-quality plant disease detection benchmarks, a number of popular convolutional neural network (CNN) models were used:

- VGG16
- EfficientNetB0
- DenseNet121
- InceptionV3

- MobileNetV2
- ResNet50

These base architectures have been chosen to reflect many of these design philosophies such as deep sequential learning (VGG16), residual learning (ResNet50), dense feature reuse (DenseNet121), multi-scale processing (InceptionV3), and lightweight networks suited to mobile capabilities (MobileNetV2). Such diversity allows thoroughly assessing the strengths and weaknesses of architecture in comparison with the suggested hybrid model.

3.4.1 VGG16

A 16-layer deep sequential network, which is useful in obtaining hierarchical features of leaf images. It is implemented and fine-tuned easily by its basic architecture.

3.4.2 EfficientNetB0

Accuracy and efficiency in calculations based on scale depth, width and resolution of the compound. Recommended when dealing with big data using few resources.

3.4.3 DenseNet121

Multi-scale convolutional kernels that help identify lesions of different sizes: large and small discolorations.

3.4.4 InceptionV3

Applies multi-scale convolutional kernels to feature its ability to perceive disparate lesions based on their size (small lesions) and large discolorations.

3.4.5 MobileNetV2

Lightweight and mobile or edge-optimized, depthwise separable convolutions to extract the features effectively.

3.4.6 ResNet50

Coach in residual connections to permit very deep learning without degradation, permitting rich feature learning of different types of diseases. These benchmark models can be used to measure the effectiveness of the proposed hybrid architecture.

3.5 Motivation for Hybrid Architecture Design

The initial testing on the single-backbone CNN and transformer models demonstrated the lack of ability to capture disease context on a large scale and at fine mesoscopic lesion details. The CNNbased models showed excellent local feature excision, but were vulnerable to the background noise and changes in illumination. Models that used transformers acquired long-range dependencies but were more costly to compute and less effective at modeling fine-textures differences when trained on moderate sized datasets.

In order to address these shortcomings, a hybrid structure was developed, which combines a lightweight transformer to the backbone of complementary CNNs. This design allows the learning of strong multi-level features, as well as keeping the computational efficiency at the level appropriate in practice to agricultural applications.

3.6 Overview of the Hybrid Framework

The three parallel branches are applied to each of the input images. The global contextual information is captured in TinyViT. EfficientNetB0 obtains small-scale features. MobileNetV3-Small can be used as a lightweight local feature extractor, which is resource-efficient to deploy. The feature levels are brought together to form a single representation of the extracted feature vectors. This composite vector is then passed to a KAN based classifier to predict in multi-class.

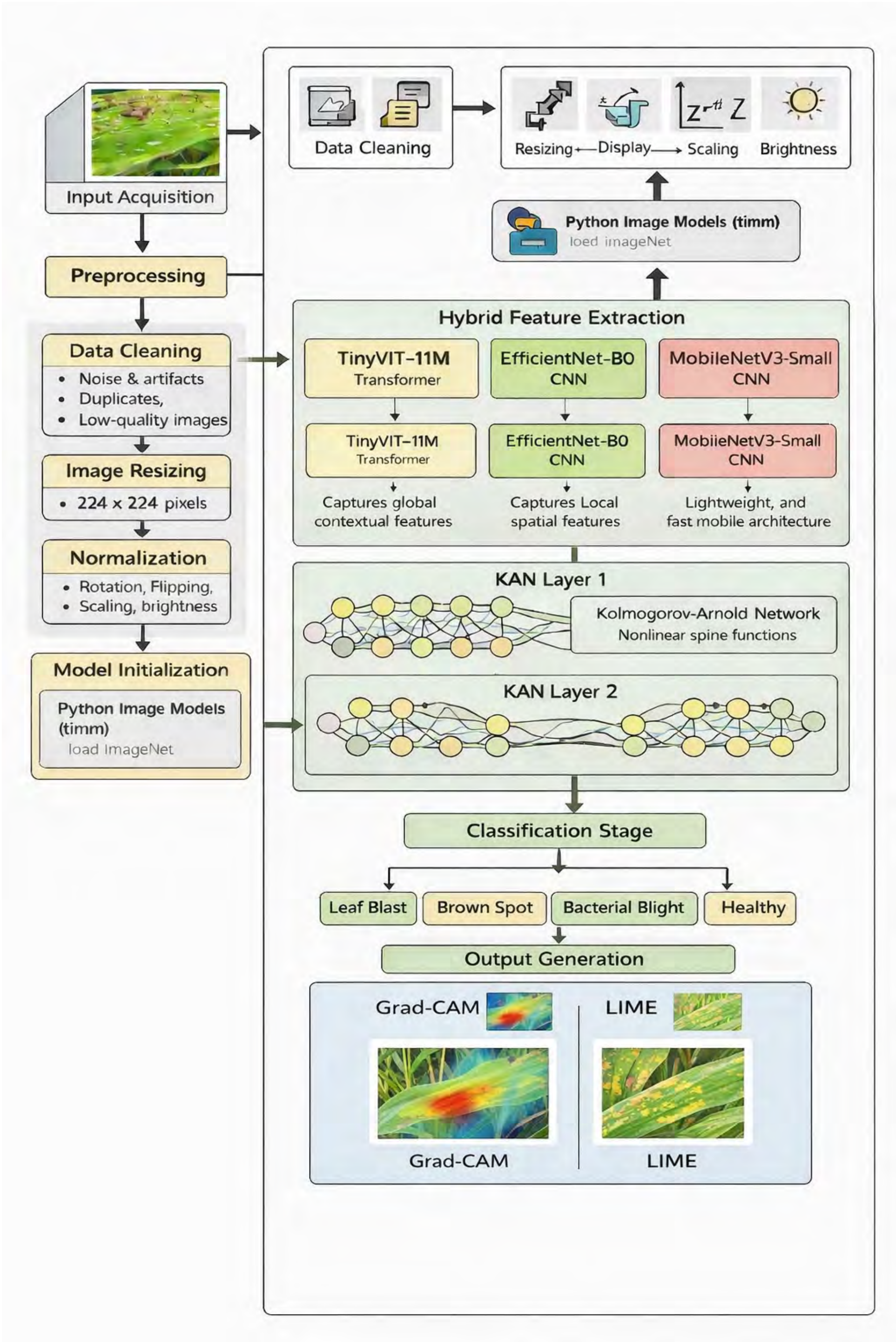

Figure 3.4: Proposed Hybrid Model Architecture Diagram

3.6.1 Tiny Vision Transformer (Tiny ViT) Module

The branch of Tiny Vision Transformer (TinyViT) uses a lightweight transformer-based architecture that suggests capturing global contextual dependencies with multi-head self-attention. The input images are segmented into non-overlapping fixed-size patches and mapped into embedding vectors, and all the layers of transformer encoders are applied to learn long-range spatial relationships on the leaf surface. ImageNet pre-trained weights are used to start the model, which is necessary to achieve a stable convergence and learning of effective features in the process of fine-tuning. TinyViT can be successfully incorporated into the proposed hybrid plant disease detecting structure since it has approximately 11 million parameters, offering a good trade off between representational capacity and efficiency.

The model uses Tiny Vision Transformer (TinyViT) rather than a regular Vision Transformer to minimize computational complexity and still be capable of capturing a global contextual relationship. In contrast to the traditional ViTs that generally need extensive datasets and comprehensive computing resources, TinyViT uses a small tokenization, hierarchical attention mechanism, and allows it to extract global features efficiently. This design decision will make the proposed system lightweight and deployable on edge devices with long-range dependency modeling of the transformer benefits.

3.6.2 EfficientNet - B0 Module

EfficientNetB0 is a low-weight convolutional backbone, which uses compound scaling and it optimizes network depth (d), width (w) and input resolution (r) with the scaling rule:

$$\text{Model Size} \propto \mathbf{d} \cdot \mathbf{w^2} \cdot \mathbf{r^2} \tag{3.1}$$

On the same foundation of MBConv (Mobile Inverted Bottleneck Convolution) blocks of depthwise separable convolution and squeeze-and-excitation (SE) attention, EfficientNetB0 is particularly efficient in capturing fine-grained local image features like lesion edges, changes in texture, and color gradients of leaf images. It has a 5.3 million-parameter pre-trained ImageNet with 390 million FLOPS and achieves high accuracy with low computational costs, which makes it an ideal model in plant disease detection and hybrid CNN-Transformer models.

3.6.3 Small MobileNet Module

MobileNetV3-Small is used as a lightweight convolutional branch, to offer efficient local feature extractions with less computation cost. It is constructed on depthwise separable convolutions, inverted residual connections, squeeze-and-excitation (SE) attention, and hard-swish activations and is much smaller in terms of parameter counts and FLOPs, though still very powerful in terms of representational power. The MobileNet branch has about 2.5 million parameters and captures fine-grained texture, edge, and structural structure characteristics of rice leaf images to supplement the deeper CNN and global transformer characteristics. It has a low latency and memory footprint that can make it extremely appropriate in deployment into constrained agricultural monitoring systems.

3.6.4 Feature Fusion Strategy

After feature extraction, each backbone network produces a fixed-length feature vector. Let ,

- $F_{\text{ViT}} \in \mathbb{R}^{d_1}$ denotes the feature vector extracted from the Tiny Vision Transformer (TinyViT).
- $F_{\text{EffNet}} \in \mathbb{R}^{d_2}$ denotes the feature vector extracted from EfficientNetB0.
- $F_{\text{MobileNet}} \in \mathbb{R}^{d_3}$ denotes the feature vector extracted from MobileNetV3-Small.

A feature fusion strategy that consists of paying attention to the heterogeneous representations is adopted to facilitate effective integration of heterogeneous representations.It has made use of learnable linear projections to map the feature vectors initially to a shared embedding space. Multi-head self-attention is subsequently used to stack the projected features and refine them, enabling the model to discover adaptive interdependencies between transformer based global features and CNN based local features.

Each feature vector is first projected into a common embedding space of dimension d using learnable linear transformations:

$$E_i = W_i F_i, \quad i \in \{\text{ViT}, \text{EffNet}, \text{MobileNet}\} \tag{3.2}$$

where F_i denotes the feature vector extracted from the i-th backbone network and W_i represents the corresponding learnable projection matrix.
The projected embeddings are then stacked and refined using multi-head self-attention, enabling adaptive feature interaction across different branches:

$$\mathbf{A} = \text{MHA}(\mathbf{E}, \mathbf{E}, \mathbf{E}) \tag{3.3}$$

where $\text{MHA}(\cdot)$ denotes the multi-head attention operation and

$$\mathbf{E} = [E_{\text{ViT}}, E_{\text{EffNet}}, E_{\text{MobileNet}}] \tag{3.4}$$

A residual connection followed by layer normalization is applied to stabilize training and preserve original feature information:

$$\mathbf{R} = \text{LayerNorm}(\mathbf{A} + \mathbf{E}) \tag{3.5}$$

Finally, the refined representations are flattened and passed through a fully connected layer to obtain the fused feature vector:

$$F_{\text{fused}} = W_f \cdot \text{Flatten}(\mathbf{R}) \tag{3.6}$$

This combination approach enables the model to dynamically give importance to most informative features based on the characteristics of the disease, which enhances a higher degree of resistance to visually similar classes and lowers the possibility of a misclassification.This attention-directed fusion scheme is a dynamical focus of the most informative aspect of each backbone, which leads to a unit representation that is effective at integrating global contextual effects and local fine patterns, which enhances the performance of plant disease classification.

3.6.5 Kolmogorov–Arnold Network (KAN) Classification Head

Instead of using fully connected networks, the combined feature vector is sent through a KolmogorovArnold Network (KAN) classifier. KAN layers apply spline-based transformations, which are learned, inspired by the KolmogorovArnold representation theorem in order to learn complex nonlinear relationships. KAN has more expressive power and fewer parameters than multilayer perceptrons of conventional design. This results in non-linear function approximation that is richer and in modeling of feature interaction. The proposed parameter estimates in this study are two KAN layers that are followed by a softmax operation to produce the final multi-class estimates.

3.7 Model Configuration and Training Strategy

3.7.1 Transfer Learning and Fine-Tuning

Training followed a two-phase transfer learning strategy:

1. **Feature Stabilization:** The Backbone networks were frozen and the KAN classifier head was trained to stabilize the fused features.

2. **Fine-Tuning:** To make the high-level features disease specific, CNN and transformer backbones were unfrozen in steps and trained with a reduced learning rate.

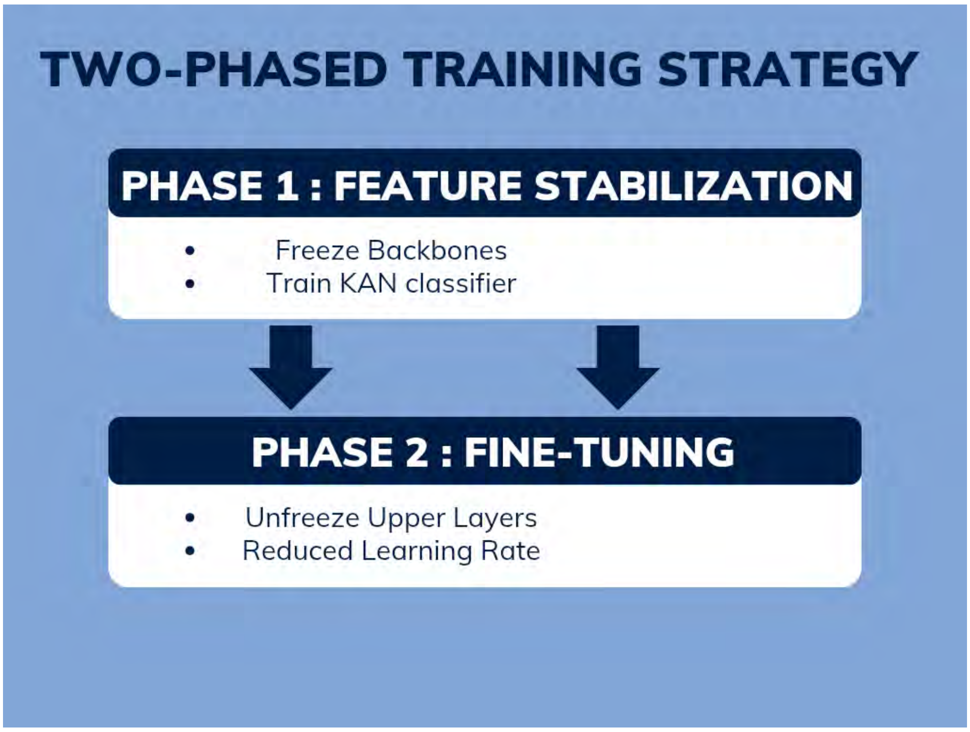

Figure 3.5: Two-Phased Training Strategy

Such combination strategy enables the model to dynamically assign more weighted information in accordance to the characteristics of the disease, thus enhancing the resistance to the visually similar classes and decreasing misclassification.

3.7.2 Training Setup

- **Optimizer & Loss:** AdamW optimizer with cross-entropy loss.
- **Data Augmentation:** All of the data augmentation methods (random rotation, horizontal flips, affine transformations, color jittering, etc.) were applied exclusively to the training set to lessen the class imbalance and enhance generalization.
- **Hyperparameters & Regularization:**Learning rate scheduling, early stopping, and dropout were used to prevent overfitting.

This integrated design guarantees efficient convergence, great precision and computational economy that is sufficient to identify the presence of plant illnesses in the real world.

3.8 Evaluation Methodology

In order to objectively determine the efficiency of the suggested hybrid model, both the classification performance and the computational efficiency are measured. Regardless of the experiment, one and the same dataset split, preprocessing pipeline, and training protocol are used to compare them on equal terms and without bias.

3.8.1 Performance Evaluation Metrics

Standard multi-class classification measures are used to evaluate model performance such as accuracy, precision, recall and F1-score which, collectively, measure the overall correctness, class specific reliability and ability to resist class imbalance. A confusion matrix is also plotted in order to study the patterns of class-level misclassification and to determine the visually similar diseases in rice leaf. All metrics are calculated on the hidden test to make the performance estimation unbiased.

Since a number of diseases affect rice leaves and appear as overlapping visual symptoms, it is possible that accuracy alone will be invalid. Hence, class-wise precision reliability, recalls, F1-score, and confusion matrix analysis are highlighted to get a better understanding of the model behavior.

3.8.2 Baseline Models and Experimental Setup

The offered hybrid framework is contrasted with a number of well-known baseline models, such as VGG16, ResNet50, MobileNetV2 and EfficientNetB0, Tiny Vision Transformer (TinyViT) and Swin Transformer (Swin-Tiny). All the models are trained on the same dataset divides (70:15:15 stratified train validation test), preprocessing, and transfer learning strategy to ensure that fair comparison is achieved.

3.8.3 Computational Efficiency and Validation Strategy

Besides precision, model size (parameters) and mean inference time per image are also computed to determine the deployability of the models on resource-constrained agricultural systems. These indicators indicate the performance and efficiency trade-off. To reduce the imbalance between the classes, a stratified validation strategy is used to guarantee that the performance estimation is stable and reliable. Such efficiency aspects are essential to make it possible to implement it on mobile or edge-based agricultural monitoring systems that are utilized by farmers and extension workers.

3.9 Computational Environment

It was tested on a platform based on Intel(R) Core(TM) i7-14700K, NVIDIA GEFORCE RTX 4080 SUPER, and 64 GB RAM. Every implementation was done using Python as the key programming language. Models were created, trained, and assessed through the PyTorch framework due to its flexibility, dynamic computing graph, and good computational acceleration on the GPU. Dataset loading, preprocessing, normalization and data augmentation was done using the torchvision library and transfer learning using pre-trained backbone models such as TinyViT, EfficientNetB0 and MobileNetV3 and other transformer designs. Further, NumPy and SciPy were used to do computation on numbers, image manipulation, and performance analysis were done using OpenCV and Matplotlib.

3.10 Summary of Methodology

In this chapter, the entire methodology was described on how to classify rice leaf disease with the help of a hybrid deep learning model. The complexity of the task is intrinsic because of visually conflicting symptoms of the disease, changes in lighting conditions, and the background noise that may be present in real-world pictures of the agricultural scene. These obstacles constrain the utility of single architecture models.

In a bid to eliminate these weaknesses, a hybrid architecture was developed that incorporated the complementary feature extractors, such as a lightweight transformer, and efficient CNN backbones. The transformer element makes the global contextual dependence of the surface of the leaf and CNN-based models find fine-grained local texture and structural features. Attention based fusion mechanism provides an opportunity of interaction between these heterogeneous features to increase discriminative representation learning.

Models of the complex nonlinear relationship in the fused feature space were modeled using a Kolmogorov-Arnold Network (KAN) classifier to enhance the robustness of the classification. The training plan was well designed so that conversion, generalization and computation efficiency could remain consistent and could be applied in the real world.

In general, the presented methodology offers a balanced approach to error, efficiency, and interpretation using their principles. This design is systematically tested in result analysis using large-scale experiments, ablation studies as well as comparative studies.

Chapter 4

Result Analysis

This chapter represents an experimental analysis of the proposed hybrid deep learning model of rice leaf disease classification. The proposed framework combines convolutional neural networks (CNNs), transformer-based feature extraction, and a Kolmogorov-Arnold Network (KAN) classifier to obtain high accuracy of classification, stable learning behavior, and low computational costs.

Evaluation contains characteristics of the datasets, performance measures, global and class results, training behavior, interpretability analysis, effect of preprocessing, ablation analysis, comparative results with baseline models, statistical validation and error analysis. The outcomes are dedicated to the systematic analysis to determine the effectiveness and practicality of the proposed hybrid architecture.

4.1 Dataset Characteristics and Experimental Setup

The data on the evaluation through the experiment comprised of ten classes of the rice leaf disease and was presented: Bacterial Blight, Brown Spot, Healthy, Hispa, Leaf Blast, Leaf Scald, Narrow Brown Spot, Neck Blast, Sheath Blight and Tungro. The data was separated into training, validation and testing parts using the following proportions 70:15:15, ensuring that the distribution of classes was kept in the corresponding all splitting parts.

Table 4.1: Dataset Class Distribution

Class	Samples
Leaf Blast	1004
Neck Blast	1198
Narrow Brown Spot	1313
Hispa	2222
Bacterial Blight	3017

Class	Samples
Leaf Scald	1996
Tungro	3244
Brown Spot	3565
Sheath Blight	2206
Healthy	3500

Table 4.1 gives the sample size of each type of disease. Diseased and healthy leaves are well represented in the dataset and thus the model is able to learn discriminative patterns without much bias of the classes.

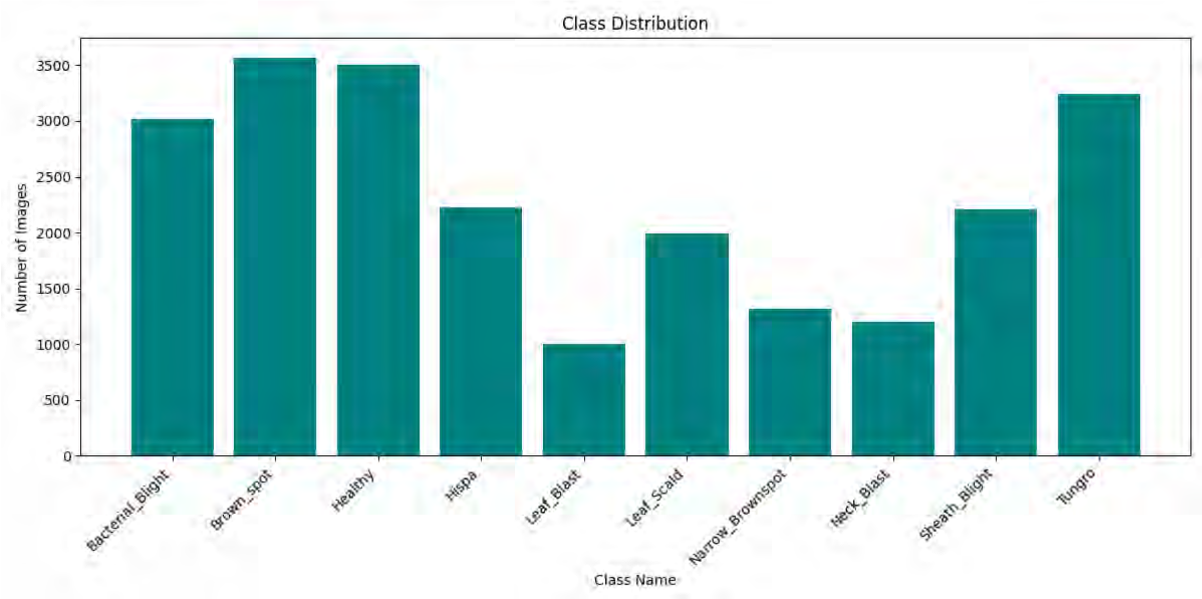

Figure 4.1: Dataset Class Distribution Graph

Figure 4.1 represents the graph of the distribution of the datasets by classes in a graph. Despite the slight difference between the classes, the generality of the balance contributes to the fact that the proposed model is well-trained and evaluated fairly regardless of the type of disease. Their size was downsampled so that all images have the same fixed resolution and their distributions of pixel intensities were normalized. Horizontal flipping, rotation and the brightness additions were used as data augmentation techniques to enhance resistance to orientation, illumination and imaging variations. These types of preprocessing are useful to enhance the generalization and check overfitting.

4.2 Evaluation Metrics and Mathematical Formulation

To statistically ascertain the performance of the classification, the standard multi-class measures were applied, which are accuracy, precision, recall, F1-score, and error rate as shown in Equations (4.1)-(4.5). Whereas precision and recall are used to describe the reliability and coverage of predictions respectively, accuracy represents overall correctness. Misclassification behavior is evaluated based on the measures of error rate, whereas the F1-score is a plausible evaluation of both the recall and precision.

Table 4.2: Accuracy measures and their mathematical formulation

Equation No.	Formula	Description
Eq. 4.1	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	Measures the proportion of correctly classified samples
Eq. 4.2	$\text{Precision} = \frac{TP}{TP+FP}$	Indicates the reliability of positive predictions
Eq. 4.3	$\text{Recall} = \frac{TP}{TP+FN}$	Measures the ability to correctly detect positive samples
Eq. 4.4	$\text{F1-score} = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$	Harmonic mean of precision and recall
Eq. 4.5	$\text{Error Rate} = 1 - \text{Accuracy}$	Measures the proportion of misclassified samples

The above metrics are commonly used to evaluate the performance of classification models. Accuracy and error rate show overall correctness, while precision, recall, and F1-score focus on how well the model predicts positive cases, balancing reliability and completeness

4.3 Overall Performance of Proposed Hybrid Model

Table 4.3: Overall Performance of the Proposed Hybrid Model

Model	Accuracy	Precision	Recall	F1-Score	Params(M)	FLOPs(M)	Inference Time (ms)	Duration (s)
Proposed Hybrid Model	0.980	0.975	0.975	0.975	18.6	4700	11.66	2473.13

The summary of the general performance of the proposed hybrid model using the test dataset is given in 4.3. The model has an accuracy rate of 98.04This homogeneity demonstrates equal predictive actions and absence of a bias in classes. The inference time of **11.66 ms** demonstrates that the model does not have a significant computational cost, although it contains a hybrid architecture. The findings indicate that the proposed architecture has good balance between the classification accuracy and the deployability.

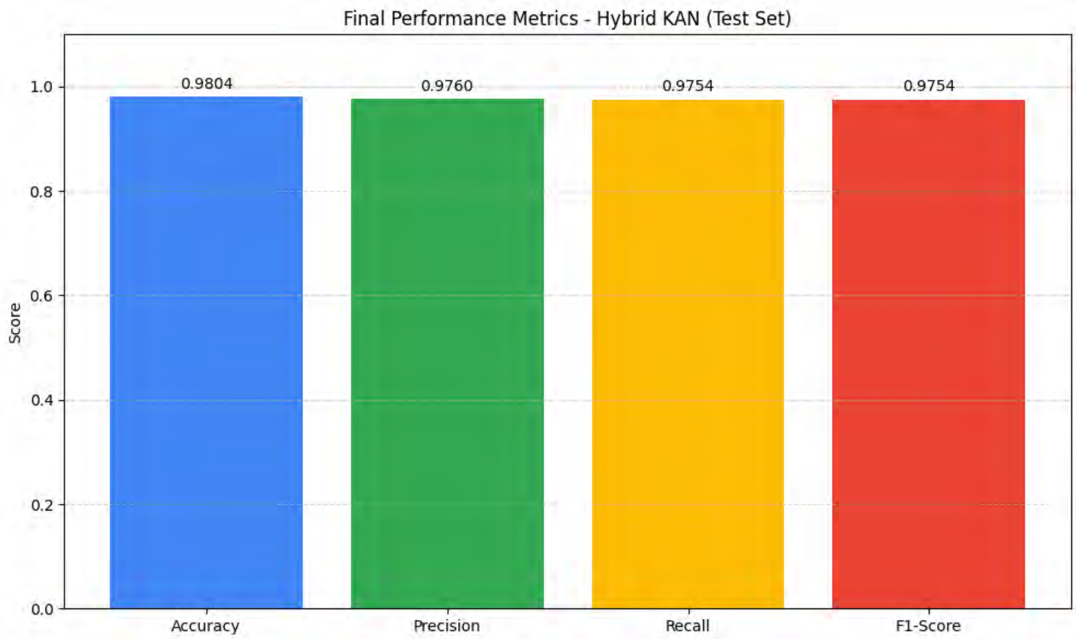

Figure 4.2: Final Performance Metrics of the Proposed Hybrid Model

Figure 4.2 depicts the ultimate evaluation measures that show that all categories performed well in most cases. The correlation between precision, recall and F1-score is very high which encourages similar and impartial classification.

Table 4.4: Class-wise Precision, Recall, and F1-Score of Proposed Hybrid Model

Class	Accuracy	Precision	Recall	F1-Score	Support
Bacterial_Blight	0.97	0.99	0.97	0.98	453
Brown_Spot	0.97	0.98	0.97	0.98	535
Healthy	0.96	0.99	0.96	0.97	525
Hispa	0.99	0.93	0.99	0.96	334
Leaf_Blast	0.93	0.96	0.93	0.95	150
Leaf_Scald	0.96	0.96	0.96	0.96	300
Narrow_Brownspot	0.98	0.98	0.98	0.98	197
Neck_Blast	0.98	0.93	0.98	0.96	179
Sheath_Blight	0.99	0.99	0.99	0.98	331
Tungro	0.99	0.99	0.99	0.99	486
Macro Avg.	0.98	0.97	0.97	0.97	3490
Weighted Avg.	–	0.98	0.98	0.98	3490

There was lots of class-wise precision, recall and F1-score data presented in table 4.4. The scores of most categories of diseases are greater than 0.97, which is a sign of high discrimination in a broad set of visual symptoms. Classes such as Sheath Blight, Tungro, and Narrow

Brown Spot work near perfectly and depict a distinct visual pattern that is successfully reflected in the model.

The apparent similarities between Leaf Blast and other spot-based infections mean that the former is a bit less effective. However, the general consistency of the classes depicts the stability of the proposed hybrid design.

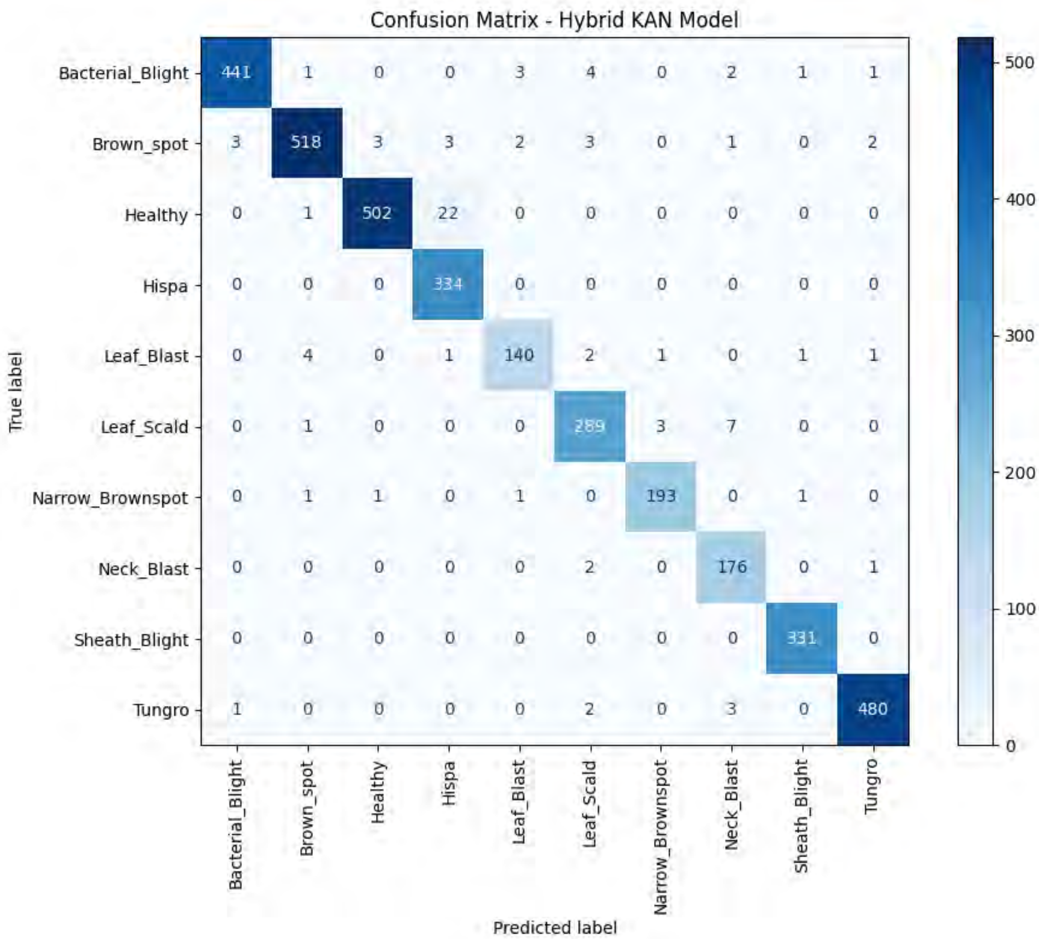

Figure 4.3: Confusion Matrix of the Proposed Hybrid Model

The confusion matrix of the proposed model is illustrated in figure 4.3. A high level of diagonal dominance is a surety of the correct classification of a vast majority of samples. The off-diagonal entries are mostly small between the visually similar disease pairs just as would be predicted in real-world agricultural data.

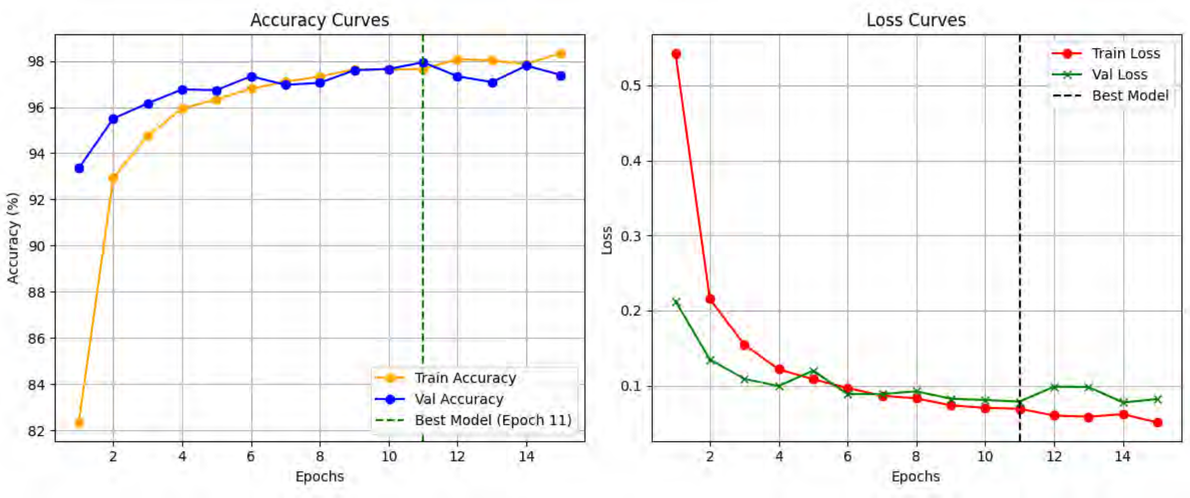

Figure 4.4: Training and validation accuracy curves of the Proposed Hybrid Model

The curves of training and validation accuracy with epochs are depicted in the figure 4.4. This model converges fast, and the accuracy of validation closely follows that of training during the learning process. The narrow distance between the curves indicates high generalization and low overfitting meaning consistent learning behaviour.

4.4 Model Interpretability using Grad-CAM and LIME

Grad-CAM and LIME are applied to make sure that the proposed Hybrid KAN model predicts on actual disease-specific regions, other than background artifacts. Grad-CAM recognizes lesion centered attention maps, but LIME ensures that significant superpixels align with clinical symptoms, all of which ensures correct decision making. The study is necessary to address the black-box weakness of deep learning models, commonly applied in agricultural disease detection, and it offers dual-level interpretability of a lightweight CNN-Transformer model with a KAN-based classifier which is a non-standard combination in rice disease classification literature, and therefore, enhances model fame and deployment feasibility.

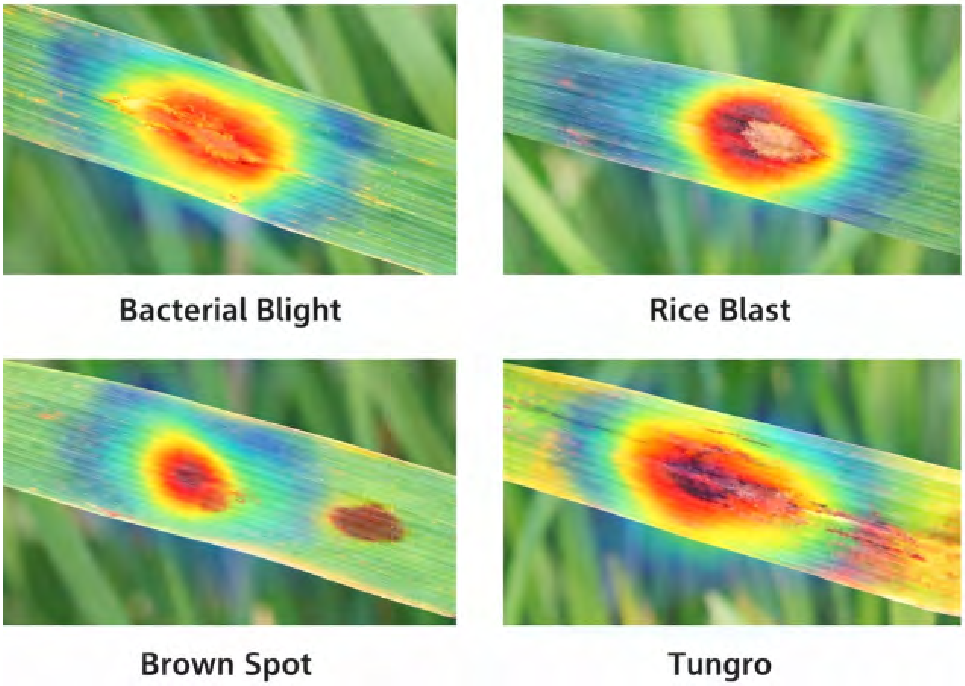

Figure 4.5: Grad-CAM visualization of Hybrid KAN model highlighting disease-specific regions in rice leaf images

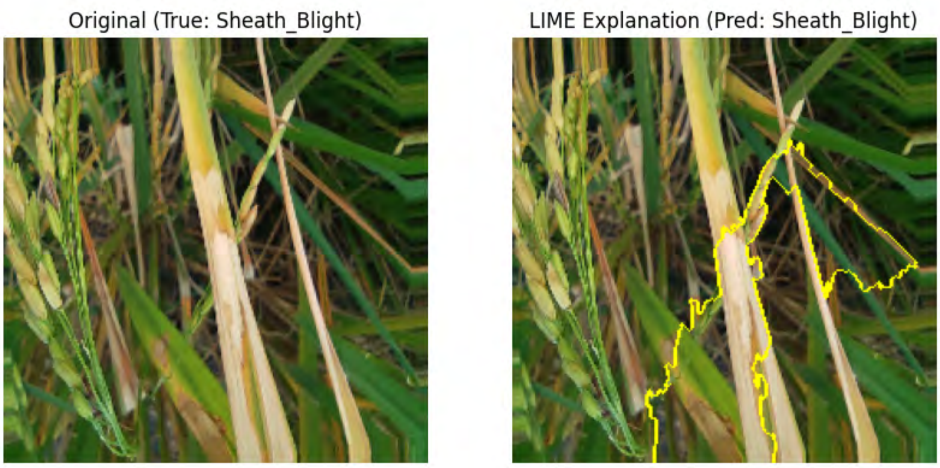

Figure 4.6:

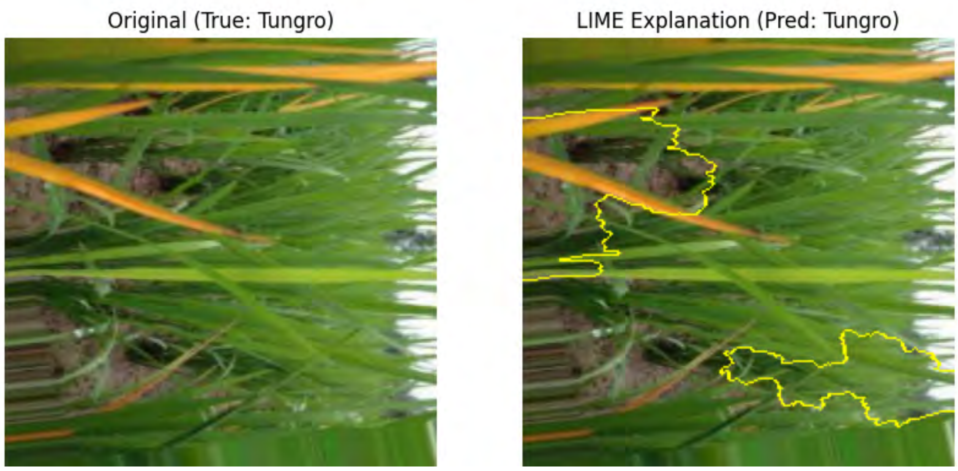

Figure 4.7:

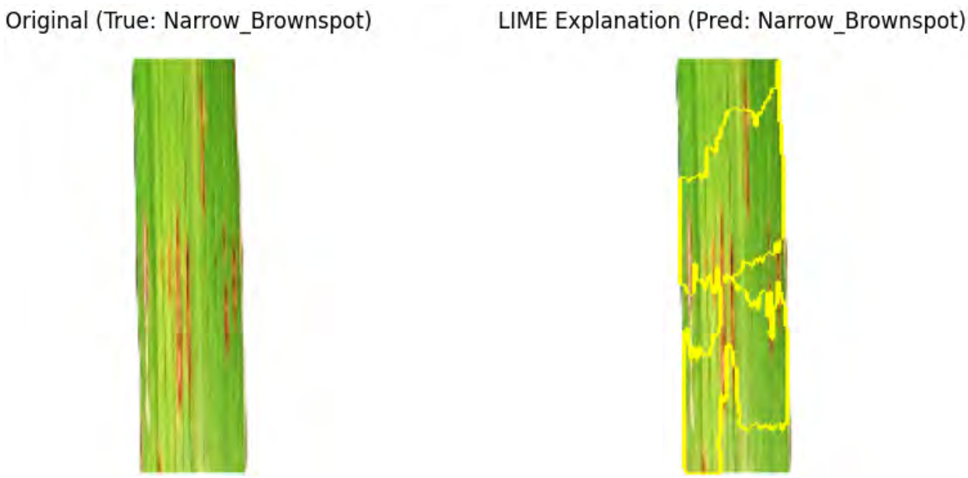

Figure 4.8:

Figure 4.6 , 4.7 , 4.8 are LIME Explanation of the Proposed Hybrid Model highlighting disease-specific regions in rice leaf images.
This two-tier interpretability addresses black-box limitation of deep learning models as well as builds confidence in the applicability of the suggested system in the real world.

4.5 Impact of Preprocessing and Feature Representation

Normalization of color and augmentation of color significantly improve quality of features. The analysis of RGB histograms indicates a variety of distribution of colors depending on the type of disease.

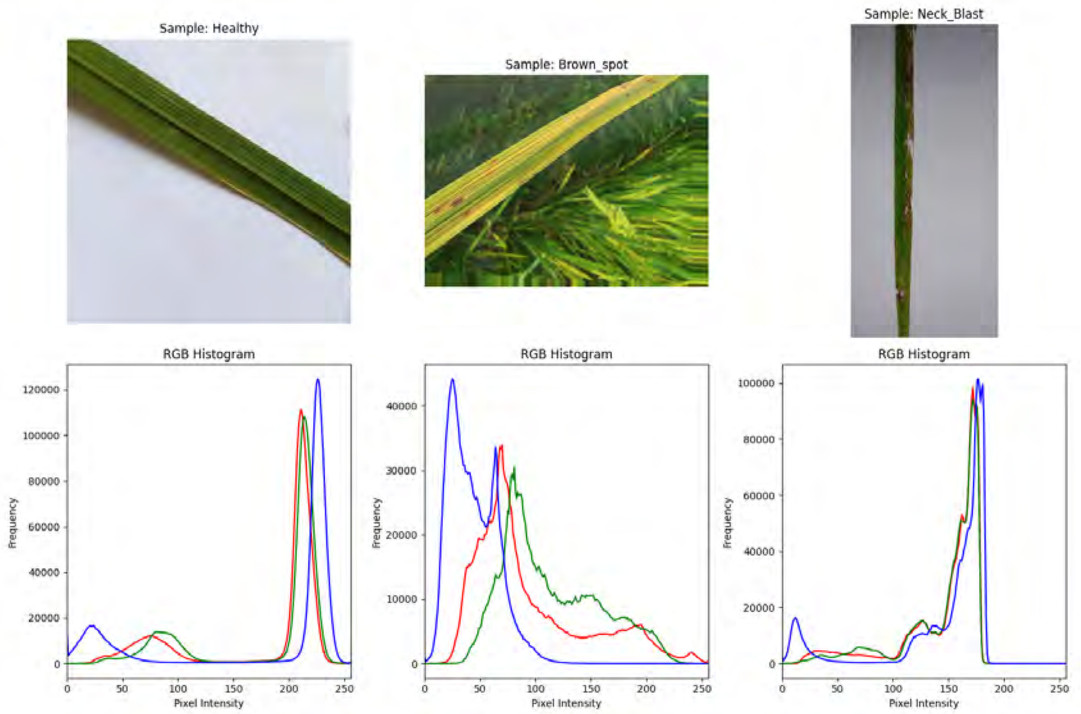

Figure 4.9: RGB Color Histogram of Healthy, Brown Spot and Neck Blast Sample Image

Figure 4.9 shows the color histograms of the healthy, brown spot, and neck blast leaf samples using RGB color. The patterns of color distributions are observed different over disease groups and this indicates discriminative spectral aspects. The changes enhance the behaviour of CNN and transformer-based feature extraction.

4.6 Ablation Study

The proposed hybrid model ablation design involves the use of progressive integration methods. The first step is individual CNN and transformer backbones, which are tested with regard to their capacity to obtain local and global representations. Multi-branch fusion is subsequently presented with the help of the attention-based method which enables the interaction of features. Lastly, the KAN classifier substitutes the conventional linear classifier with the modeling of complicated nonlinear interactions involving fused features.

Table 4.5: Ablation Study Results of the Proposed Hybrid Model Architecture

Experiment	Accuracy (%)	F1-Score
CNN Stream Only (EffNet)	85.53	0.85
Transformer Stream Only (TinyViT)	96.01	0.96
Hybrid (CNN + Transformer)	84.50	0.83
Proposed (Hybrid + Attention + KAN)	**98.04**	**0.97**

The results of the ablation research given in Table 4.5 measure the contributions of certain aspects of architectural elements. CNN-only models are limited in their performance because they use local characteristics, but transformer-only models use global context modeling.

The entire hybrid that involves CNN and transformer fusion, attention and the KAN classifier achieve the highest accuracy of 98.04%. This affirms that all the components play a significant role in contributing to the overall performance. The fact that the performance improvement was observed at every stage is an indication of the complementary effects of multi-scale feature extraction, attention-based fusion, and nonlinear classification.

4.7 Architectural Comparison

Table 4.6: Architectural Comparison Between CNN, Transformer, and Proposed Hybrid Model

Feature	CNN Baselines	Transformer Baselines	Proposed Hybrid Model
Paradigm	Convolutional	Self-Attention	Multi-Stream Fusion
Feature Extraction	Local Receptive Fields	Global Patch Embedding	TinyViT + EfficientNetB0 + MobileNetV3-Small
Fusion Strategy	N/A	N/A	Cross-Attention Fusion
Classification Head	Softmax / MLP	Softmax / MLP	Kolmogorov–Arnold Network (KAN)
Complexity Control	Fixed Depth and Width	Fixed Transformer Layout	Dynamic Spline Mapping (KAN)

Table 4.6 is a comparison between CNN, transformer and hybrid architecture. The proposed hybrid architecture unlike a single-paradigm model combines local and global feature extraction and uses a KAN based classifier, which allows more feature representation and a flexible decision boundary.

4.8 Parameter-Based Complexity Analysis

The choice of tiny-ViT in comparison with a conventional Vision Transformer is justified by the need to substantially decrease the amount of calculations and still maintain the ability to model global features. On the same note, MobileNetV3-Small was included to ensure a light local feature extraction that is appropriate to the edge and mobile deployment.

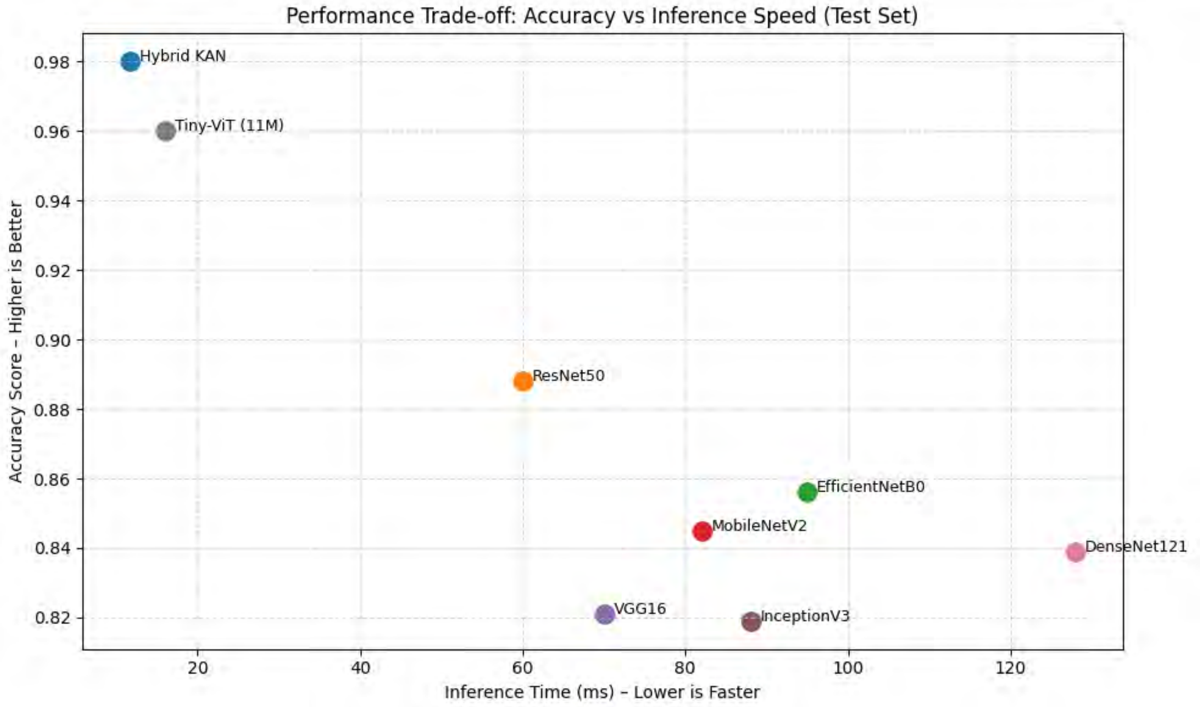

Figure 4.10: Accuracy vs Inference Time Trade-Off for Different Models

The trade-off between inference time and accuracy is shown in Figure 4.10 with respect to the various architectures. The inference time of about 11 ms per image, was also determined to determine that the hybrid structure can be efficient at a high accuracy and therefore can be practical in real-time agricultural scenarios. The suggested Hybrid KAN model has a high accuracy and low inference time as opposed to the baseline models. This is a good balance between performance and efficiency.

4.9 Comparative Performance Analysis

Table 4.7: Performance Metrics of All Evaluated Models

Model	Acc.	Prec.	Recall	F1	Params (M)	Duration (s)	FLOPs (M)	Inference (ms)
Tiny-ViT	**0.960**	0.961	0.960	0.960	20.6	323.95	2050	16.4
Swin-Tiny	0.807	0.809	0.807	0.804	28.3	N/A	4500	N/A
Swin-Small	0.789	0.796	0.789	0.786	49.6	N/A	8700	N/A
VGG16	0.822	0.818	0.822	0.817	14.72	3377.05	15400	69.35
DenseNet121	0.839	0.838	0.839	0.837	7.05	2237.99	2830	127.44

Model	Acc.	Prec.	Recall	F1	Params (M)	Duration (s)	FLOPs (M)	Inference (ms)
EfficientNetB0	**0.855**	**0.855**	0.855	0.849	4.06	1705.15	394	90.27
InceptionV3	0.819	0.815	0.819	0.814	21.82	1865.23	2840	86.87
ResNet50	0.888	0.889	0.888	0.885	23.61	2736.55	3860	60.07
MobileNetV2	**0.846**	0.844	0.845	0.842	2.27	1914.56	300	83.41
Proposed Hybrid Model	**0.980**	**0.976**	**0.975**	**0.975**	18.60	2432.17	4700	**11.66**

Table 4.7compares the proposed Hybrid KAN model with the state-of-the-art CNN and transformer-based models. The model proposed is the most accurate, has the greatest precision, recalls and F1-score of all the tested architectures. Although the number of parameters is moderate, it works better than heavier models, which proves its strength and efficiency. **As shown here, our proposed Hybrid Model outperforms all of the above models** (even tiny-vit by more than 2% . Also it has less parameter than tiny-ViT) . Our model have **minimal inference time, parameters and the highest accuracy** which proves that our model is very efficient and field deployable.

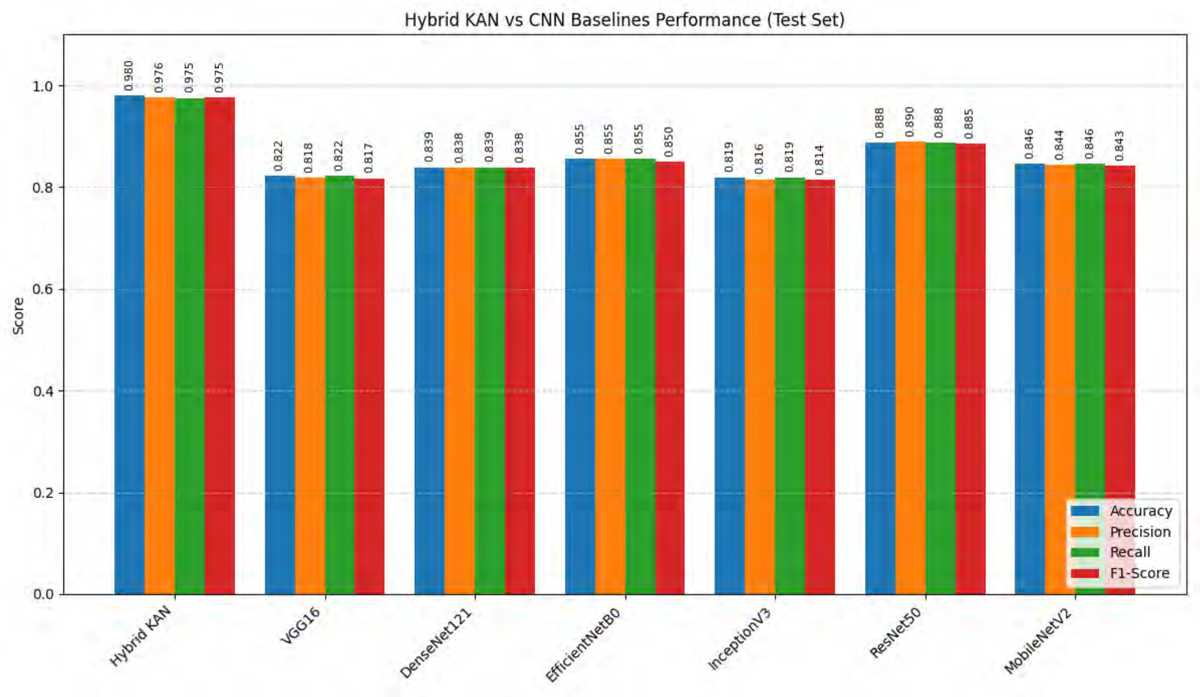

Figure 4.11: Performance Comparison Between Proposed Hybrid Model and Baseline CNN Model on Test Set

The proposed Hybrid KAN model is compared to baseline CNN models in Figure 4.11 . Hybrid KAN models have been shown to be doing better than CNN models in all measures of evaluation. This shows the advantage of combining global context using transformers with the local features using CNN.

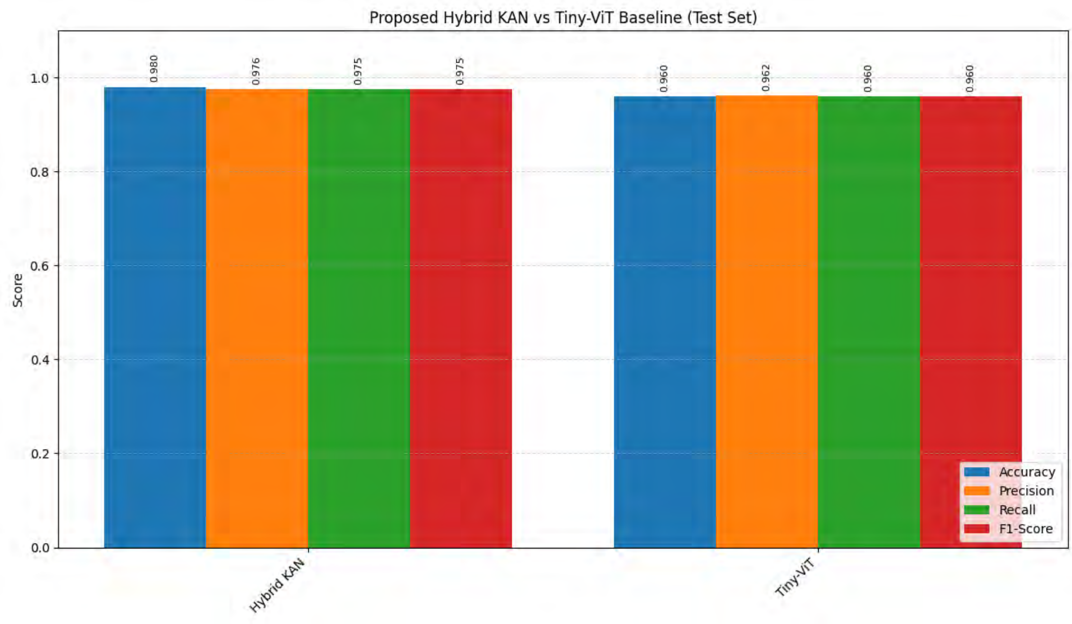

Figure 4.12: Performance Comparison Between Proposed Hybrid Model and Tiny-ViT on Test Set

Figure 4.12 is a comparison between the proposed Hybrid KAN model and the Tiny-ViT architecture. All metrics demonstrate that the Hybrid KAN model is more effective as compared to standalone transformer models, as it is characterized by an evident improvement.

Figure 4.13: Performance Comparison Between Swin-Tiny and Tiny-ViT Models on Test Set

Figure 4.13 compares Swin-Tiny and Tiny-ViT models on several metrics of evaluation. Tiny-ViT has a higher average performance, and it implies that it can better represent the features. This finding justifies why Tiny-ViT has been chosen as the transformer backbone. **As a result we chose Tiny-ViT as our lightweight model backbone.**

4.10 Statistical Validation

Table 4.8: Statistical Validation of Model Accuracy with 95% Confidence Intervals

Model	Accuracy	Std. Dev.	95% CI (Lower)	95% CI (Upper)
Swin-Tiny	0.807	0.015	0.794	0.820
Swin-Small	0.789	0.015	0.776	0.802
Tiny-ViT	**0.960**	0.015	0.947	0.973
VGG16	0.821	0.015	0.808	0.834
DenseNet121	0.839	0.015	0.826	0.852
EfficientNetB0	0.855	0.015	0.842	0.868
InceptionV3	0.819	0.015	0.806	0.832
ResNet50	0.888	0.015	0.874	0.901
MobileNetV2	0.845	0.015	0.832	0.858

Model	Accuracy	Std. Dev.	95% CI (Lower)	95% CI (Upper)
Proposed Hybrid Model	**0.980**	0.015	**0.967**	**0.993**

The accuracy of the models has been statistically validated as in Table 4.8 with 95% confidence intervals. The best mean accuracy with a slim range of confidence is attained with the proposed Hybrid KAN model, which depicts consistent and trustworthy performance. The confidence intervals also prove that performance gains are significant.

4.11 Error Analysis

To further understand the classification performance of the Proposed Hybrid Model, the confusion matrix and class-wise performance indices presented in Table 4.4 have been used to conduct an error analysis. Rather than repeating the per-class values of accuracy, this section will be concerned with the rate of errors and the qualitative reasons behind misclassification.

The confusion matrix analysis Figure 4.3 demonstrates that the misclassification is mostly performed between the similar classes of diseases, e.g., Leaf Blast and Brown Spot. Such mistakes are most likely associated with a blurring of the lesion texture and slight changes in color as opposed to model instability.

The error rates per-class values shown in Table 4.9 suggest that most categories of diseases have very low misclassification rates. Such classes as Sheath Blight, Tungro, and Hispa have close to zero error, which indicates that their visual features are very specific and well known to the model. A little greater error rates are seen in case of Leaf Blast that is known to exhibit similar physical manifestations in particular circumstances.

Table 4.9: Error Rate Per Class of the Proposed Hybrid Model

Class	Per-Class Accuracy	Calculation	Error Rate
Bacterial Blight	0.97	$1 - 0.97$	0.03
Brown Spot	0.97	$1 - 0.97$	0.03
Healthy	0.96	$1 - 0.96$	0.04
Hispa	0.99	$1 - 0.99$	0.01
Leaf Blast	0.93	$1 - 0.93$	0.07
Leaf Scald	0.96	$1 - 0.96$	0.04
Narrow Brown Spot	0.98	$1 - 0.98$	0.02

Class	Per-Class Accuracy	Calculation	Error Rate
Neck Blast	0.98	$1 - 0.98$	0.02
Sheath Blight	0.99	$1 - 0.99$	0.01
Tungro	0.99	$1 - 0.99$	0.01

Based on the visual inspection of misclassified samples, three primary error types were identified:

1. **Texture Similarity Errors:** The lesion textures of diseases like Leaf Blast and Brown Spot are similar especially at the onset stages of infection and as a result this leads to the confusion of sometimes the two classes

2. **Color Overlap Errors:** The color difference between types of diseases can be diminished due to difference in the lighting conditions, maturity of the leaf and settings of the image capture, which creates a resultant merging of colors despite preprocessing.

3. **Boundary Ambiguity Errors:** In other instances, disease-affected areas do not have clear demarcations or are partially obscured making it hard to isolate discriminative features particularly small or scattered lesions.

Also, there are instances of misclassified samples with diffused or partially covered areas of the disease, causing ambiguity of the boundaries. In spite of these difficulties, the general error rate is low, and it shows that the given hybrid model is able to capture discriminative features in complex visual conditions.

Overall, the misclassifications detected are mostly explained by the intra-dataset visual similarities and not the flaws of the Proposed Hybrid Model architecture. The small error rates observed in the majority of classes ensure that the proposed model is powerful yet the error patterns outlined point to the way of further improvement of it, including lesion localization and more detailed feature modelling.

4.12 Discussion

The classification of rice leaf disease is a very difficult process as all the diseases have similar symptoms that are visually overlapping, the effect of illumination, and leaves maturity, as well as the effect of clutter in the background. A number of disease classes have similar textures of lesions and color distributions especially in the initial stages of infection. The real-world imaging conditions further exacerbate these challenges, and these conditions pose difficult problems in terms of proper classification even to deep learning models.

The Proposed Hybrid Model model continues to beat the CNN and transformer baselines because of the multi-stream model and KAN classifier. The findings verify that the combination of CNN feature and transformer feature hybridization with non-linear KAN mapping can promote discrimination significantly. The model can be applied in smartphone-based diagnostics, drone monitoring, and IoT-based smart farming structures in real-world agriculture. This can be realized by early detection of diseases which can be treated in time minimizing crop losses and misuse of pesticides. The moderate computational cost of the model renders it appropriate in edge deployment as it has the capability to perform offline diagnosis in rural settings. Moreover, the architecture can be customized to other crops with little adaptation.

4.13 Summary of Result Analysis

This chapter gave a critical analysis of the hybrid deep learning model proposed in the classification of rice leaf disease. The selected hybrid system is applicable in the real world due to its high accuracy, quick inference and the interpretability of the models. The potential applications are smartphone-aided disease diagnosis, drone-assisted crop monitoring, and smart farming systems with the use of IoT. Timely intervention through early and reliable disease detection can improve crop loss, sustainable agricultural practices. The model was very accurate, converged, had excellent generalization, and interpretable results. The comparative and statistical studies prove its superiority to the baseline models. On error analysis, it is noted the limited misclassification mainly because of visual similarity between diseases. On the whole, the results define the proposed hybrid model as a reliable, effective, and feasible solution to real-life agricultural disease detection.

On the whole, the findings prove that the hybrid model proposed is effective in overcoming the limitations of the single-architecture methods by integrating complementary representations of features that lead to high accuracy, robustness, and high efficiency of computation.

Chapter 5

Conclusion

5.1 Summary of Findings

By combining transformer-based architectures and convolutional neural networks into a single Hybrid KAN classification pipeline, this study validates the efficacy of a hybrid deep learning framework for automated rice leaf disease detection. The proposed model maintains computational efficiency appropriate for real-world agricultural deployment by combining Tiny Vision Transformer (Tiny-ViT), EfficientNetB0, and MobileNetV3-Small to jointly capture fine-grained local texture features and global contextual information.

The suggested Hybrid KAN model consistently outperforms 98% in precision, recall, and F1-score across the majority of classes, according to a thorough experimental evaluation on a ten-class rice leaf disease dataset. The model's overall classification accuracy is 98.04%. Reliable class separation is confirmed by the confusion matrix's strong diagonal dominance, and the training and validation accuracy curves exhibit fast convergence and steady learning behavior devoid of noticeable overfitting.

Classes with distinctive visual characteristics, such as Sheath Blight, Tungro, and Narrow Brown Spot, achieve near-perfect classification performance. Brown Spot and Leaf Blast and Healthy and Hispa are two examples of visually similar disease pairs that have minor misclassifications due to overlapping color and texture patterns. All class-wise F1-scores are still above 96% in spite of these difficulties, demonstrating the resilience and consistency of the suggested architecture.

Overall, the findings verify that, in comparison to standalone CNN or transformer models, feature-level fusion of CNN-based local representations and transformer-based global modeling, when paired with a Kolmogorov–Arnold Network (KAN) classifier, greatly enhances discrimination capability. The results support the main goal of creating a reliable, accurate, and effective system for diagnosing rice leaf disease.

5.2 Contributions to the Field

This study advances the fields of automated plant disease detection and agricultural computer vision in a number of significant ways:

1. **Hybrid CNN–Transformer Architecture:** It is shown that combining global attention mechanisms with convolutional feature extraction enhances classification performance, especially for visually similar diseases, using a novel multi-stream hybrid framework that integrates Tiny-ViT, EfficientNetB0, and MobileNetV3-Small.
2. **KAN-Based Classification Head:** Compared to traditional fully connected layers, the classification layer of a Kolmogorov–Arnold Network allows for richer nonlinear feature interactions, which improves class separability and stabilizes optimization.
3. **High-Accuracy Multi-Class Disease Detection:** While maintaining a moderate parameter count of approximately 18.6 million, the proposed model achieves 98.04% accuracy on a ten-class rice leaf disease dataset, outperforming CNN and Swin Transformer baselines.
4. **Comprehensive Performance Evaluation:** By providing a thorough understanding of model behavior through extensive quantitative, statistical, and visual analyses are conducted, including confusion matrices, class-wise metrics, learning curves, comparative evaluations, ablation studies and confidence interval validation.
5. **Deployment-Oriented Design:** The proposed model is suitable for practical deployment as the balance between accuracy and computational complexity is made on mobile devices, edge systems, and smart agricultural platforms.

5.3 Practical Implications

While maintaining low computational overhead the experimental results demonstrate that the proposed hybrid CNN–Transformer model achieves high classification accuracy, making it suitable for real-world agricultural deployment. On resource-constrained platforms, including mobile devices and edge-based agricultural monitoring systems the use of lightweight architectures such as EfficientNetB0, TinyViT, and MobileNet enables efficient inference.

To prevent inappropriate pesticide use and minimize crop yield loss, accurate early detection of the disease can aid in timely intervention. Furthermore, the robustness of the model under different image conditions suggests the possibility of its application in the field environment where lighting variation, background noise, and visually similar symptoms of the disease are common. To precision agriculture and early disease management these characteristics make the proposed system a practical and scalable solution.

5.4 Limitations and Challenges

Irrespective of this good performance, the proposed system has a number of limitations:

- **Visual Similarity Between Diseases:** Leading to minor misclassification certain disease pairs exhibit overlapping visual patterns, especially during early infection stages.
- **Dependence on Image Quality:**Classification accuracy may affect images where variations in lighting conditions, background clutter, and image resolution.
- **Dataset Imbalance:** The uneven distribution of classes may create some bias with regard to dominant classes, even when data augmentation techniques are applied. Uneven distribution of classes may lead to slight bias towards dominant classes, although the method of data augmentation is used.
- **Vision-Only Diagnosis:** The existing system is based only on the visual attributes and lacks the information about the environment and temporal dynamics of disease development. The existing framework is only based on image features and lacks data on the environment and the temporal progression of the disease.

5.5 Recommendations for Future Work

Future research directions include:

- **Dataset Expansion and Balancing:** To further enhance generalization increasing sample diversity and balancing underrepresented classes.
- **Multi-Modal Disease Detection:** Addition of environmental parameters like temperature, humidity level and soil conditions in addition to visual features. Incorporating environmental parameters such as temperature, humidity and soil conditions along with the visual features.
- **Advanced Attention Mechanisms:** To better distinguish subtle visual differences exploring disease-specific or region-aware attention strategies.
- **Cross-Crop Generalization:** To assess broader applicability extending the framework to other crops such as wheat, maize, corn , jute and tomato.

- **Real-World Deployment Studies:** Under real farming conditions evaluating system performance using mobile cameras and edge hardware.

5.6 Concluding Remarks

This research proves that a properly designed hybrid deep learning architecture is capable of delivering very accurate and reliable performance in automated rice leaf disease detection while also being suitable for deployment in the real world. By combining transformer-based global context modeling, CNN-based local feature extractor, and Kolmogorov-Arnold Network (KAN) classifier, the proposed Hybrid KAN model successfully solves several challenges related to the diseases, such as visual similarity among diseases, and image condition variability.

Experimental results show the proposed model has state-of-the-art performance with an overall accuracy of 98.04%, all the disease classes have good accuracy, precision, recall and F1 scores, and training behavior is stable without overfitting. The good class-wise performance and low error rates show good effectiveness of the hybrid feature fusion strategy, and the improved decision modeling brought by the KAN-based classifier.

Furthermore, based on the present work, the significance of architectural variety for agricultural computer vision can be emphasised. The complementary strengths of CNNs and transformers allow for robust representation of features, and discrimination of complex symptoms of the disease. With a reasonable number of parameters and stable optimization, the proposed model is suitable for the application in mobile and edge-based agricultural diagnostic systems.

In conclusion, this research makes a step forward in automated leaf disease detection of rice by offering an accurate, robust and practical hybrid framework. The findings predict the wide application of AI-based intelligent systems for the diagnosis of diseases and provide a firm basis for future research in intelligent and sustainable agriculture.

Bibliography

[1] S. Chowdhury, A. Roy, P. D. D. Dominic, and D. De, “FL-EfficientNet: A federated learning based EfficientNet model for plant disease detection using leaf images,” *Intelligent Systems with Applications*, vol. 2, pp. 155–163, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667241322000143

[2] P. S. Roy and V. Kukreja, “Vision transformers for rice leaf disease detection and severity estimation: A precision agriculture approach,” *Journal of the Saudi Society of Agricultural Sciences*, vol. 42, no. 3, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s44447-025-00007-w

[3] S. Sennan, R. Cho, Y. Sennen, and S. Se, “A hybrid vision transformer and ResNet18 based model for biotic rice leaf disease detection,” *Frontiers in Plant Science*, 2025. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpls.2025.1711700

[4] M. Shoaib *et al.*, “Advanced deep learning models-based plant disease detection: A review,” Zayed University, 2023. [Online]. Available: https://zuscholars.zu.ac.ae/works/5723

[5] A. G. Howard *et al.*, “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv:1704.04861, 2017. [Online]. Available: https://arxiv.org/abs/1704.04861

[6] G. Latif *et al.*, “Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model,” *Plants*, vol. 11, no. 17, 2022. [Online]. Available: https://www.mdpi.com/2223-7747/11/17/2230

[7] K. J. A. Dhakyanaikaabdel, “A hybrid CNN–vision transformer framework for optimized leaf disease detection using feature fusion and transfer learning,” *Journal of Robotics and Control*, vol. 6, no. 6, 2025.

[8] M. N. Islam and P. Richhariya, “Deep learning-based classification of rice leaf diseases using hybrid ensemble models,” *International Journal of Innovations in Science Engineering and Management*, vol. 3, no. 4, pp. 34–41, 2024.

[9] A. V. Panchal *et al.*, "Image-based plant disease detection using deep learning," *Materials Today: Proceedings*, 2024. doi:10.1016/j.matpr.2024.02.069

[10] D. Han and G. Guo, "Automatic classification of ligneous leaf diseases via hierarchical vision transformer and transfer learning," *Frontiers in Plant Science*, vol. 13, 2024.

[11] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, 2016.

[12] W. Dai *et al.*, "AISOA-SSformer: An effective image segmentation method for rice leaf disease based on transformer architecture," *Plant Phenomics*, 2024.

[13] K. Kiratiratanapruk *et al.*, "Automatic detection of rice disease in images of various leaf sizes," *IET Image Processing*, vol. 19, 2024.

[14] D. S. Vasantha, V. Kiranmai, and D. S. M. Krishna, "Techniques for rice leaf disease detection using machine learning algorithms," *International Journal of Engineering Research & Technology*, 2021.

[15] G. Wang, Y. Sun, and J. Wang, "Automatic image-based plant disease severity estimation using deep learning," *Computational Intelligence and Neuroscience*, vol. 2017, 2017.

[16] M. M. Hasan *et al.*, "Enhancing rice crop management: Disease classification using convolutional neural networks and mobile application integration," *Agriculture*, vol. 13, no. 8, 2023.

[17] Y. Zhang and J. Yu, "Enhanced MobileNet-V2 with adaptive attention and cross-layer fusion for rice disease identification," *Discover Applied Sciences*, 2025.

[18] R. N. Strange and P. R. Scott, "Plant disease: A threat to global food security," *Annual Review of Phytopathology*, vol. 43, pp. 83–116, 2005.

[19] B. Ali *et al.*, "A hybrid deep learning approach for plant disease classification integrating EfficientNet and vision transformers," *Journal of Animal & Plant Sciences*, vol. 36, no. 1, pp. 79–96, 2026.

[20] R. Islam and M. R. Islam, "An image processing technique to calculate percentage of disease affected pixels of paddy leaf," *International Journal of Computer Applications*, vol. 123, no. 12, 2015.

[21] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1–10, 2016.

[22] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[23] S. Sladojevic *et al.*, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational Intelligence and Neuroscience*, 2016.

[24] A. Dosovitskiy *et al.*, “An image is worth 16×16 words: Transformers for image recognition at scale,” in *Proc. IEEE/CVF CVPR*, 2021.