A Comparative Variability Analysis of Pre-Trained Deep Learning Models for Sugarcane Leaf Disease Classification

Bhaliya Nirali Nitinbhai[1*], and Bijal Talati[1]

[1]Computer Science and Engineering Department, Parul University, Vadodara, Gujarat, India.
*Corresponding Author: Bhaliya Nirali Nitinbhai. Email: nirali.bhaliya270184@paruluniversity.ac.in

_____

Abstract: This reseach provides a variability study of well known pre-trained deep learning models in automated sugarcane leaf disease classification. VGG16, VGG19, ResNet50, InceptionV3, MobileNetV2, DenseNet121 and EfficientNetB0 were used to transfer learning in various training conditions, such as different epochs and learning rates. The experimental findings indicate that EfficientNetB0 provided the highest performance regularly with a highest accuracy, precision, recall, and F1-score of 0.89 at 50 epochs with a learning rate of 0.001, meaning that it extracts features better and also uses fewer parameters. ResNet50 and DenseNet121 were also found to be stable with accuracies that lie between 0.82 and 0.85 and therefore can be used in practical applications. Nevertheless, InceptionV3 demonstrated relatively low accuracy. The effect of using increasing epochs over time past 50 was a small overfitting in certain models, whereas there was little improvement in learning rates. On the whole, EfficientNetB0 can be considered the most appropriate model, although a tailored architecture can be even more effective.

Keywords: Sugarcane Leaf; Disease Classification; Transfer Learning; Pre-trained Models; Variability Analysis

1. Introduction

Sugarcane is one of the most economically important commercial crops, playing a vital role in global sugar and bio-ethanol production. However, sugarcane cultivation is severely affected by various foliar diseases that significantly reduce crop yield and quality if not detected at an early stage. Traditional disease identification methods rely heavily on visual inspection by agricultural experts, which is time-consuming, subjective, and often impractical for large-scale farming. Recent advancements in computer vision and deep learning have enabled automated plant disease diagnosis using leaf images, offering a promising solution for early detection and precision agriculture. Transfer learning using pre-trained deep learning models has gained considerable attention due to its ability to leverage knowledge from large-scale datasets and adapt it to domain-specific agricultural applications with limited labeled data.

Convolutional Neural Networks have demonstrated remarkable performance in image classification tasks, including plant disease recognition. Pre-trained architectures such as VGG, ResNet, Inception, MobileNet, DenseNet, and EfficientNet have been widely adopted due to their proven feature extraction capabilities and computational efficiency. However, the performance of these models can vary significantly depending on dataset characteristics, network depth, parameter optimization, and training configurations such as learning rate and number of epochs. Despite extensive use of transfer learning in agricultural image analysis, there remains a lack of systematic studies that analyze the variability and robustness of multiple pre-trained models under identical experimental conditions for sugarcane leaf disease classification. Understanding these performance variations is essential for selecting reliable models for real-world deployment and for identifying limitations of generic architectures when applied to crop-specific disease patterns.

Motivated by these challenges, this paper presents a comprehensive variability analysis of multiple pre-trained deep learning models for sugarcane leaf disease classification. The study evaluates VGG16, VGG19, ResNet50, InceptionV3, MobileNetV2, DenseNet121, and EfficientNetB0 across different training epochs and learning rates to assess their classification accuracy, precision, recall, and F1-score. The experimental findings reveal that EfficientNetB0 consistently achieves superior performance, while ResNet50 and DenseNet121 demonstrate robust and stable results, whereas other models exhibit limitations such as overfitting or poor generalization. Furthermore, the observed performance gaps highlight the necessity of developing a customized deep learning architecture tailored specifically to sugarcane disease characteristics. The insights provided in this work aim to guide researchers and practitioners in selecting suitable deep learning models and inspire future research toward more robust and crop-specific disease classification frameworks.

1.1. Related Works

Recent studies have demonstrated the effectiveness of deep learning and transfer learning techniques for sugarcane leaf disease detection and classification. Islam et al. [1] utilized a custom Convolutional Neural Network (CNN) and ResNet50 architecture to identify sugarcane leaf diseases, showing that deep feature extraction significantly outperforms traditional approaches. Comprehensive reviews and benchmarking studies have further highlighted the growing adoption of transfer learning–based CNN models in plant leaf disease classification across multiple crops [2], [3]. Beyond disease identification, image-based deep learning frameworks have also been applied to related agricultural tasks such as nutrient content estimation using modified ResNet architectures [4]. Comparative analyses of deep learning models for early sugarcane disease detection have reported promising results using architectures such as VGG, ResNet, and DenseNet [5], while optimization-driven and hybrid deep learning approaches have also been explored to improve classification accuracy [6]. Broad surveys on transfer learning in agriculture emphasize its suitability for handling limited labeled datasets commonly encountered in real-field scenarios [7], and several studies have demonstrated successful sugarcane disease classification using deep neural network frameworks [8].

More recent research has focused on enhancing model robustness, scalability, and deployment feasibility. Edge–cloud frameworks integrating deep learning and transfer learning have been proposed for plant disease detection using remote sensing data [9], while predictive systems combining environmental and crop characteristics have also been explored [10]. Transfer learning has been successfully applied to multi-class leaf disease classification across various crops, including date palm and hops, indicating its generalizability beyond sugarcane [11], [12]. EfficientNet-based architectures have gained attention for their parameter efficiency and strong performance in sugarcane leaf disease detection [13]. Ensemble learning strategies combining multiple fine-tuned transfer learning models have further improved classification reliability [14]. Additionally, mobile and application-oriented solutions integrating deep learning for real-time sugarcane disease diagnosis have been proposed [15], [16]. Broader reviews and datasets related to sugarcane health monitoring, including satellite-based and image-based approaches, have contributed to advancing research in this domain [17], [18], while comparative studies across plant disease detection techniques continue to validate the superiority of deep learning–based solutions [19].

Despite the substantial progress reported in literature, several limitations remain. Some studies rely on shallow architectures such as AlexNet [20] or focus on explainable artificial intelligence models without extensive comparative evaluation across multiple architectures [21]. While hybrid CNN–Transformer models and attention-based networks have shown promise for sugarcane leaf disease diagnosis [22], [23], they often demand large-scale datasets and high computational resources, limiting their applicability to moderately sized, real-field datasets. Furthermore, although data augmentation has been shown to improve classification performance [24], many existing works do not systematically analyze model variability under identical experimental settings, including learning rate selection, epoch optimization, and augmentation impact. Most importantly, most current approaches employ generic pre-trained architectures originally designed for natural image classification, which may not optimally capture sugarcane-specific disease characteristics. These limitations highlight the need for a comprehensive

variability analysis of pre-trained models and motivate the development of customized deep learning architectures tailored specifically for sugarcane leaf disease classification.

2. Materials and Methods

Figure 1 describes the dataset used for sugarcane leaf disease classification, the preprocessing and augmentation strategies applied, and the detailed working principles of the selected pre-trained deep learning models. A comprehensive experimental framework was designed to ensure fair comparison among all models under identical training conditions.
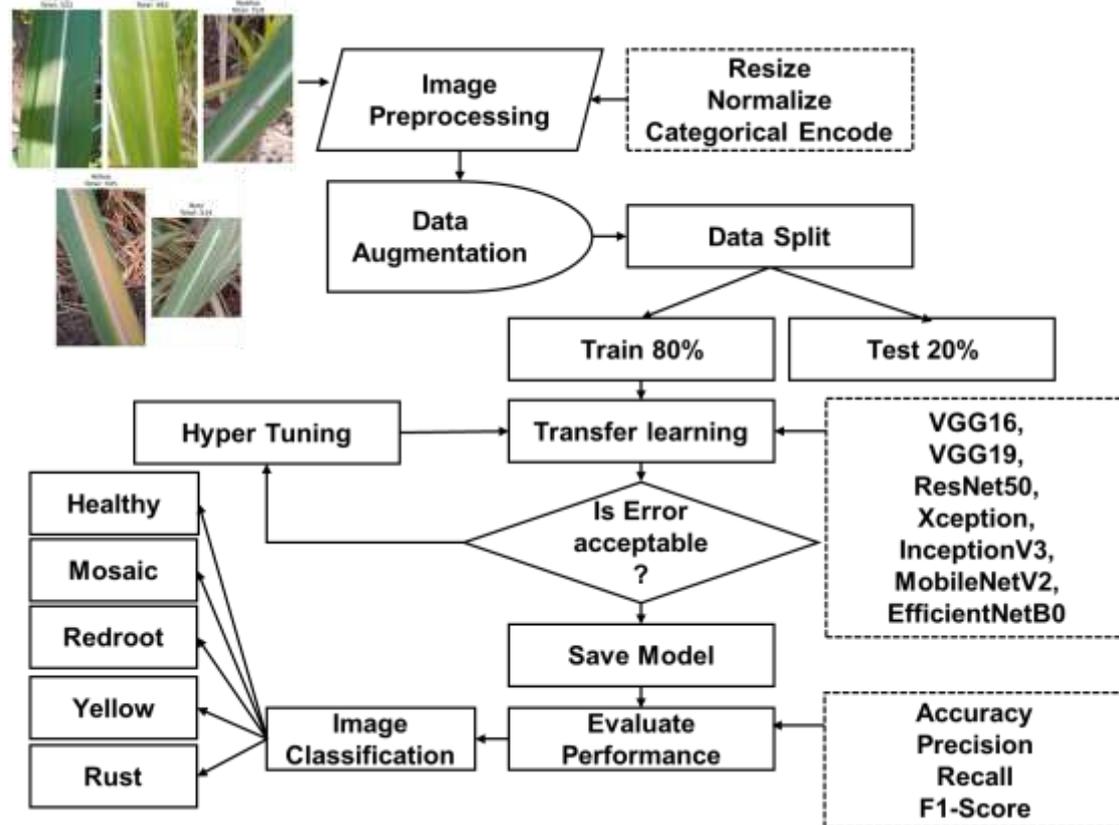

Figure 1. Proposed System Process Diagram

2.1. Dataset Description

The dataset used in this study is a manually collected image dataset of sugarcane leaf diseases, publicly available through the Mendeley Data repository. The dataset comprises 2,521 RGB images of sugarcane leaves captured under real-field conditions using handheld devices. The images exhibit variations in illumination, background clutter, orientation, and disease severity, thereby closely representing real-world agricultural scenarios.

The dataset consists of five distinct categories, representing common sugarcane leaf conditions, including healthy leaves and four different disease classes. These categories collectively capture both visual texture variations and colour distortions associated with disease infection. The dataset is moderately sized, making it suitable for transfer learning approaches where pre-trained models are fine-tuned rather than trained from scratch.

All images were resized to model-specific input dimensions prior to training. The dataset was first split into training (80%), validation (10%), and testing (10%) subsets using stratified sampling. Data augmentation and class balancing were applied strictly after the split and exclusively to the training set to prevent data leakage.

2.2. Image Pre-processing

Prior to model training, all images underwent a standardized pre-processing pipeline. Images were resized according to the input requirements of each architecture (e.g., 224×224 for VGG, ResNet, DenseNet, EfficientNetB0, and InceptionV3). Pixel values were normalized using model-specific pre-processing

functions to align with the statistical distribution of the ImageNet dataset, on which the pre-trained weights were originally learned.

Noise removal was not explicitly applied, as convolutional neural networks are inherently capable of learning robust features from noisy inputs. However, normalization ensured faster convergence and stable gradient flow during training. No handcrafted feature extraction was used; all models learned discriminative features directly from raw pixel data.

2.3. Data Augmentation Strategy

To evaluate the impact of data variability on model generalization, all experiments were conducted with and without data augmentation. Data augmentation was applied only to the training set to prevent data leakage.

The augmentation techniques included random horizontal and vertical flipping, rotation, zooming, width and height shifting, and minor brightness adjustments. These transformations simulate real-world variations such as leaf orientation changes, camera angle differences, and illumination inconsistencies commonly observed in agricultural fields. Augmentation was particularly important given the relatively limited dataset size, as it increases effective sample diversity and reduces overfitting.

For the non-augmented experiments, the models were trained solely on the original images to analyse baseline performance and susceptibility to overfitting. Comparing augmented and non-augmented results enabled a deeper understanding of each model's robustness and generalization capability.

2.4. Transfer Learning Framework

All experiments employed a **transfer learning approach**, leveraging pre-trained weights learned from the ImageNet dataset. The convolutional base of each model was initially frozen to retain generic visual feature representations, while custom fully connected layers were appended for sugarcane disease classification. During later training stages, selective fine-tuning was applied to deeper layers to adapt model representations to domain-specific features such as lesion texture, colour patterns, and vein distortions.

The learning rate values (0.01, 0.001, and 0.0001) were selected based on commonly adopted fine-tuning practices in transfer learning literature. The upper bound (0.01) evaluates aggressive optimization behaviour, while 0.0001 assesses conservative weight updates. The epoch limit of 50 was chosen based on preliminary convergence analysis showing performance plateau beyond this point.

2.5. VGG16 and VGG19

VGG16 and VGG19 are deep convolutional neural network architectures that follow a highly uniform and straightforward design philosophy, making them widely adopted as baseline models in image classification tasks. Both architectures are built using sequential stacks of small 3×3 convolutional filters followed by max-pooling layers, enabling gradual abstraction of visual features. VGG16 consists of 16 trainable layers, while VGG19 extends this structure to 19 layers, allowing deeper hierarchical feature learning. The simplicity and consistency of the VGG architecture make it effective in capturing spatial patterns, edges, and textures in natural images.

In the context of sugarcane leaf disease classification, VGG16 and VGG19 were fine-tuned using transfer learning to adapt their pre-trained ImageNet weights to the five disease categories. The early convolutional layers successfully extracted low-level features such as edges, veins, and colour gradients, which are common across plant leaf images. Deeper layers learned more complex disease-specific patterns, including lesion shapes, discoloration regions, and texture irregularities. However, due to the large number of trainable parameters, both VGG models exhibited high computational complexity and increased memory consumption. This characteristic made them more susceptible to overfitting, particularly when trained on the non-augmented dataset. The comparison between augmented and non-augmented training revealed that data augmentation played a critical role in improving generalization, although overfitting tendencies were still observed at higher training epochs.

2.6. ResNet50

ResNet50 is a deep convolutional architecture that introduces residual learning through identity-based skip connections. These residual connections allow gradients to propagate directly across layers, effectively addressing the vanishing gradient problem that often arises in very deep networks. The

architecture comprises 50 layers organized into residual blocks, where each block learns a residual function instead of a direct mapping, enabling deeper feature extraction without degradation in performance.

For sugarcane leaf disease classification, ResNet50 demonstrated strong learning stability and robust performance. The residual structure enabled the model to capture complex disease characteristics such as subtle texture variations and elongated lesion patterns while maintaining consistent convergence behavior. Fine-tuning of the deeper residual layers allowed effective adaptation to disease-specific features without compromising the general visual representations learned during pre-training. Compared to VGG-based models, ResNet50 showed reduced overfitting and more stable validation performance across both augmented and non-augmented datasets, highlighting its robustness and suitability for medium-sized agricultural datasets.

2.7. InceptionV3

InceptionV3 employs a more complex architectural design based on parallel convolutional operations with varying kernel sizes within the same layer. This multi-branch structure enables simultaneous extraction of fine-grained local features and broader global patterns, making the model theoretically well-suited for diverse image representations. Additionally, factorized convolutions and dimensionality reduction techniques are incorporated to improve computational efficiency.

Despite these advantages, InceptionV3 exhibited limited effectiveness in the sugarcane leaf disease classification task. The architectural complexity of the inception modules typically requires large-scale datasets to fully exploit their representational power. Given the moderate size of the sugarcane dataset, the model struggled to consistently learn discriminative disease features. Even with data augmentation, validation performance remained unstable, indicating difficulty in generalizing disease patterns. These findings suggest that while InceptionV3 is powerful for large and diverse datasets, it may be less suitable for relatively constrained agricultural image datasets.

2.8. MobileNetV2

MobileNetV2 is a lightweight deep learning architecture designed specifically for deployment in resource-constrained environments such as mobile and edge devices. It achieves computational efficiency through the use of depth wise separable convolutions and inverted residual blocks with linear bottlenecks, significantly reducing the number of parameters and floating-point operations.

In this study, MobileNetV2 demonstrated rapid convergence and low computational overhead, making it attractive for real-time agricultural applications. However, its compact architecture limited its ability to model complex visual patterns associated with sugarcane leaf diseases. While basic disease features were learned effectively, the model struggled with subtle variations in texture and colour intensity. Data augmentation improved generalization to some extent, but MobileNetV2 remained more prone to performance degradation when trained for extended epochs, particularly under non-augmented conditions.

2.9. DenseNet121

DenseNet121 adopts a dense connectivity strategy in which each layer receives feature maps from all preceding layers through direct connections. This design ensures maximum feature reuse, improved gradient flow, and efficient parameter utilization. By concatenating feature maps across layers, DenseNet121 preserves both low-level and high-level information throughout the network.

DenseNet121 proved highly effective in capturing discriminative sugarcane disease representations. The dense connections allowed the model to leverage fine-grained texture details alongside higher-level semantic features, resulting in stable and reliable classification performance. Compared to deeper architectures with higher parameter counts, DenseNet121 achieved strong generalization with moderate computational cost. Its consistent performance across different learning rates, epoch settings, and augmentation strategies highlights its suitability as a robust alternative for agricultural disease classification tasks.

2.10. EfficientNetB0

EfficientNetB0 is based on a compound scaling strategy that uniformly scales network depth, width, and input resolution using a principled optimization approach. This balanced scaling enables

EfficientNetB0 to achieve high accuracy with significantly fewer parameters compared to traditional convolutional neural networks.

In this work, EfficientNetB0 consistently outperformed all other evaluated models across all experimental settings. Its well-balanced architecture effectively captured disease-specific features such as color discoloration, lesion distribution, and vein distortions while minimizing overfitting. Data augmentation further enhanced its generalization capability, although strong performance was maintained even without augmentation. EfficientNetB0 also exhibited superior stability across different learning rates and training epochs, making it the most suitable and reliable model for sugarcane leaf disease classification in this study.

2.11. Evaluation Metrics

The performance of the proposed classification framework was quantitatively evaluated using four widely adopted metrics: accuracy, precision, recall, and F1-score, which together provide a comprehensive assessment of classification effectiveness. These metrics are particularly critical in agricultural disease detection tasks, where incorrect classification especially false negatives can lead to delayed intervention and significant crop yield loss. All evaluations were performed on an unseen test dataset to ensure unbiased and reliable performance assessment.

Accuracy measures the overall correctness of the classification model and is defined as the ratio of correctly classified samples to the total number of samples. It is expressed as:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad [1]$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Precision evaluates the reliability of positive predictions by measuring the proportion of correctly identified positive samples among all samples predicted as positive. It is given by:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad [2]$$

Recall, also referred to as sensitivity, quantifies the model's ability to correctly identify actual positive samples, which is crucial for minimizing missed disease cases. Recall is calculated as:

$$\text{Recall} = \frac{TP}{TP+FN} \qquad [3]$$

The F1-score provides a balanced measure by computing the harmonic mean of precision and recall, particularly useful when class distributions are imbalanced. The F1-score is defined as:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad [4]$$

3. Results

The experimental setup for this study was implemented using the Google Collaboratory platform to leverage its cloud-based computational resources. All experiments were conducted on a system equipped with an NVIDIA Tesla T4 Graphics Processing Unit (GPU) with 16 GB of dedicated GPU memory, which enabled efficient training and fine-tuning of deep learning models. The implementation was carried out using the Python programming language with TensorFlow and Keras libraries. The models were trained using transfer learning with pre-trained ImageNet weights, and a uniform training configuration was maintained across all experiments to ensure fair comparison. The Adam optimizer was employed with learning rates of 0.01, 0.001, and 0.0001, and the categorical cross-entropy loss function was used for multi-class classification. Training was performed for up to 50 epochs with a fixed batch size, and early stopping was applied based on validation performance to mitigate overfitting. Both augmented and non-augmented datasets were evaluated under identical experimental conditions. All performance metrics were computed on an unseen test set, ensuring unbiased and reproducible evaluation of the pre-trained model variability for sugarcane leaf disease classification.

Figure 2 illustrates the dataset reading process by presenting one representative sample image from each sugarcane leaf class along with its original image count. The dataset consists of five categories, namely Healthy, Mosaic, RedRot, Rust, and Yellow, with noticeable variation in the number of images per class

(Healthy: 522, Mosaic: 462, RedRot: 518, Rust: 514, and Yellow: 505). The displayed samples highlight the visual diversity across disease categories, including differences in leaf texture, color discoloration, lesion patterns, and vein structure. This figure provides an initial insight into the dataset composition and reveals the inherent class imbalance present in the raw dataset, which can negatively affect model learning and bias classification results if not addressed appropriately.

Figure 2. Dataset Reading

Figure 3 demonstrates the dataset balancing process, where each sugarcane leaf disease class is uniformly represented with 700 images after augmentation. One sample image per class is shown along with the balanced image count, confirming that the dataset has been equalized across all five categories. Data augmentation techniques such as rotation, flipping, and zooming were applied exclusively to the training set to synthetically increase underrepresented classes while preserving disease-specific visual characteristics. This balancing strategy ensures that all classes contribute equally during model training, thereby reducing classification bias and improving generalization performance. The balanced dataset forms a robust foundation for fair comparison of pre-trained deep learning models and enhances the reliability of the experimental results.

Figure 3. Dataset Balancing

Figure 4 illustrates representative correctly classified and misclassified samples across all five disease categories, providing qualitative insight into model behavior. The visual examples highlight distinct texture, color variation, and lesion patterns that influence classification performance across Healthy, Mosaic, Red Rot, Rust, and yellow leaf classes.
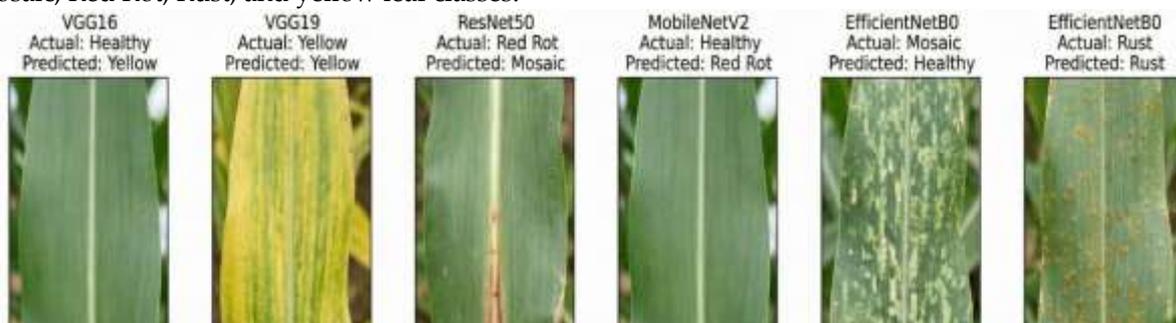

Figure 4. Sample Predictions

Table 1 presents the training and validation accuracy and loss curves for all evaluated pre-trained models under normal (non-augmented) and augmented training conditions. The plots illustrate the learning behavior of VGG16, VGG19, ResNet50, InceptionV3, MobileNetV2, DenseNet121, and EfficientNetB0 across training epochs. In the non-augmented setting, several models particularly VGG16, VGG19, and MobileNetV2 exhibit a clear divergence between training and validation curves, indicating overfitting as training progresses. In contrast, the augmented training graphs demonstrate improved stability, smoother convergence, and reduced gaps between training and validation performance,

highlighting the effectiveness of data augmentation in enhancing generalization. ResNet50 and DenseNet121 show relatively stable learning trends in both settings, while InceptionV3 exhibits slower convergence and fluctuating validation performance. EfficientNetB0 consistently achieves faster convergence with minimal overfitting, maintaining closely aligned accuracy and loss curves under both normal and augmented conditions.
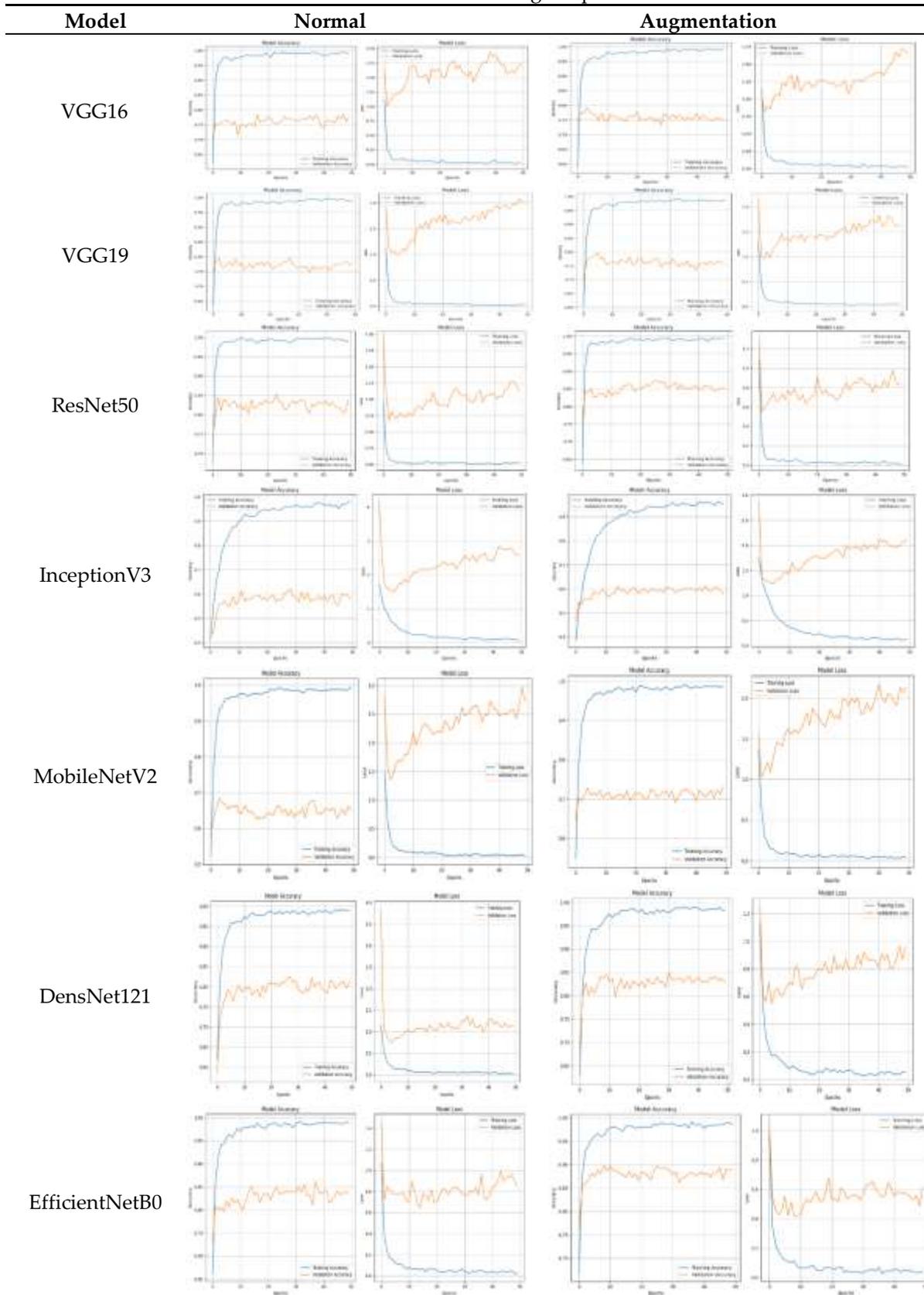
Table 1. Model Training Graphs

Model	Normal		Augmentation	
VGG16				
VGG19				
ResNet50				
InceptionV3				
MobileNetV2				
DensNet121				
EfficientNetB0				

Table 2 presents the confusion matrix analysis of all evaluated models, illustrating their class-wise prediction performance for sugarcane leaf disease classification. The results indicate that EfficientNetB0 and ResNet50 achieved better true positive rates with fewer misclassifications compared to other models.
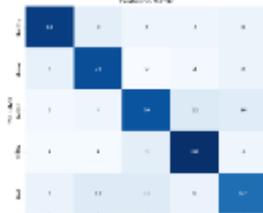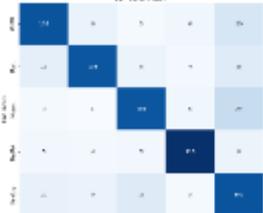
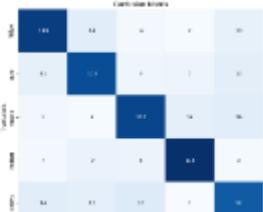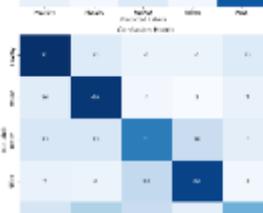**Table 2.** Models Confusion Matrix
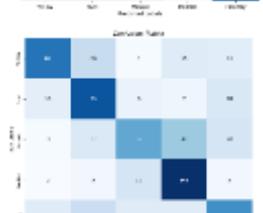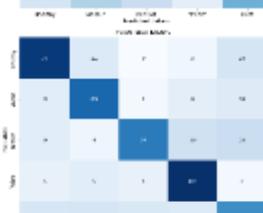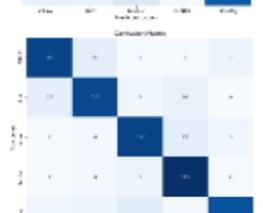
Model	Normal	Augmentation
VGG16		
VGG19		
ResNet50		
InceptionV3		
MobileNetV2		
DensNet121		
EfficientNetB0		
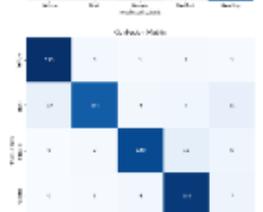
Table 3 presents the classification report of all evaluated models, summarizing their accuracy, precision, recall, and F1-score for sugarcane leaf disease classification. All reported precision, recall, and F1-scores correspond to weighted averages across classes to account for class imbalance. Additionally, macro-averaged results are reported. The findings show that EfficientNetB0 achieved the highest overall performance, followed by ResNet50 and DenseNet121, while InceptionV3 demonstrated comparatively lower metric values. Per-class precision and recall analysis indicate that Rust and Mosaic categories exhibit slightly lower recall in VGG-based models, suggesting sensitivity to subtle texture patterns.

Table 3. Models Classification Report

VGG16 — Normal

	precision	recall	f1-score	support
Healthy	0.84	0.79	0.81	105
Mosaic	0.76	0.85	0.80	92
RedRot	0.76	0.73	0.75	104
Yellow	0.79	0.85	0.82	103
Rust	0.74	0.66	0.70	101
accuracy			0.78	505
macro avg	0.78	0.78	0.78	505
weighted avg	0.78	0.78	0.77	505

VGG16 — Augmentation

	precision	recall	f1-score	support
Yellow	0.77	0.73	0.75	140
Rust	0.81	0.72	0.77	140
Mosaic	0.76	0.72	0.74	140
RedRot	0.86	0.84	0.85	140
Healthy	0.59	0.73	0.65	140
accuracy			0.75	700
macro avg	0.76	0.75	0.75	700
weighted avg	0.76	0.75	0.75	700

VGG19 — Normal

	precision	recall	f1-score	support
Healthy	0.80	0.85	0.82	105
Mosaic	0.80	0.75	0.78	92
RedRot	0.84	0.65	0.74	104
Yellow	0.80	0.83	0.82	103
Rust	0.68	0.80	0.73	101
accuracy			0.78	505
macro avg	0.78	0.78	0.78	505
weighted avg	0.78	0.78	0.78	505

VGG19 — Augmentation

	precision	recall	f1-score	support
Yellow	0.76	0.79	0.77	140
Rust	0.74	0.74	0.74	140
Mosaic	0.76	0.76	0.76	140
RedRot	0.84	0.83	0.83	140
Healthy	0.71	0.69	0.70	140
accuracy			0.76	700
macro avg	0.76	0.76	0.76	700
weighted avg	0.76	0.76	0.76	700

ResNet50 — Normal

	precision	recall	f1-score	support
Healthy	0.82	0.87	0.84	105
Mosaic	0.90	0.84	0.87	92
RedRot	0.81	0.85	0.83	104
Yellow	0.93	0.88	0.91	103
Rust	0.76	0.76	0.76	101
accuracy			0.84	505
macro avg	0.84	0.84	0.84	505
weighted avg	0.84	0.84	0.84	505

ResNet50 — Augmentation

	precision	recall	f1-score	support
Yellow	0.86	0.86	0.86	140
Rust	0.83	0.90	0.87	140
Mosaic	0.85	0.84	0.85	140
RedRot	0.86	0.90	0.88	140
Healthy	0.85	0.76	0.80	140
accuracy			0.85	700
macro avg	0.85	0.85	0.85	700
weighted avg	0.85	0.85	0.85	700

InceptionV3 — Normal

	precision	recall	f1-score	support
Healthy	0.62	0.68	0.65	105
Mosaic	0.56	0.75	0.64	92
RedRot	0.52	0.49	0.50	104
Yellow	0.65	0.64	0.65	103
Rust	0.54	0.36	0.43	101
accuracy			0.58	505
macro avg	0.58	0.58	0.57	505
weighted avg	0.58	0.58	0.57	505

InceptionV3 — Augmentation

	precision	recall	f1-score	support
Yellow	0.71	0.57	0.63	140
Rust	0.55	0.66	0.60	140
Mosaic	0.55	0.41	0.47	140
RedRot	0.58	0.77	0.66	140
Healthy	0.55	0.49	0.52	140
accuracy			0.58	700
macro avg	0.59	0.58	0.58	700
weighted avg	0.59	0.58	0.58	700

MobileNetV2 — Normal

	precision	recall	f1-score	support
Healthy	0.68	0.73	0.70	105
Mosaic	0.62	0.68	0.65	92
RedRot	0.74	0.55	0.63	104
Yellow	0.68	0.78	0.72	103
Rust	0.54	0.50	0.52	101
accuracy			0.65	505
macro avg	0.65	0.65	0.65	505
weighted avg	0.65	0.65	0.65	505

MobileNetV2 — Augmentation

	precision	recall	f1-score	support
Yellow	0.73	0.74	0.73	140
Rust	0.69	0.66	0.68	140
Mosaic	0.72	0.69	0.70	140
RedRot	0.77	0.85	0.81	140
Healthy	0.74	0.71	0.72	140
accuracy			0.73	700
macro avg	0.73	0.73	0.73	700
weighted avg	0.73	0.73	0.73	700

DensNet121 — Normal

	precision	recall	f1-score	support
Healthy	0.68	0.73	0.70	105
Mosaic	0.62	0.68	0.65	92
RedRot	0.74	0.55	0.63	104
Yellow	0.68	0.78	0.72	103
Rust	0.54	0.50	0.52	101
accuracy			0.65	505
macro avg	0.65	0.65	0.65	505
weighted avg	0.65	0.65	0.65	505

DensNet121 — Augmentation

	precision	recall	f1-score	support
Yellow	0.84	0.83	0.83	140
Rust	0.80	0.88	0.80	140
Mosaic	0.81	0.84	0.82	140
RedRot	0.82	0.89	0.85	140
Healthy	0.89	0.78	0.83	140
accuracy			0.83	700
macro avg	0.83	0.83	0.83	700
weighted avg	0.83	0.83	0.83	700

EfficientNetB0 — Normal

	precision	recall	f1-score	support
Healthy	0.90	0.90	0.90	105
Mosaic	0.77	0.83	0.80	92
RedRot	0.86	0.80	0.83	104
Yellow	0.92	0.85	0.88	103
Rust	0.74	0.88	0.77	101
accuracy			0.84	505
macro avg	0.84	0.84	0.84	505
weighted avg	0.84	0.84	0.84	505

EfficientNetB0 — Augmentation

	precision	recall	f1-score	support
Yellow	0.91	0.95	0.93	140
Rust	0.93	0.79	0.85	140
Mosaic	0.91	0.86	0.88	140
RedRot	0.86	0.91	0.89	140
Healthy	0.84	0.94	0.89	140
accuracy			0.89	700
macro avg	0.89	0.89	0.89	700
weighted avg	0.89	0.89	0.89	700

Table 4 presents a comparative performance analysis of all evaluated pre-trained models under non-augmented and augmented training conditions. The results indicate that data augmentation significantly improves the performance of most architectures, particularly DenseNet121 and MobileNetV2, which show

substantial gains in accuracy and F1-score after augmentation. EfficientNetB0 achieves the best overall performance, reaching accuracy, precision, recall, and F1-score of 0.89 with augmentation, confirming its superior generalization capability. ResNet50 also benefits marginally from augmentation, demonstrating stable and high performance across all metrics. In contrast, VGG16 and VGG19 exhibit slight performance degradation with augmentation, suggesting sensitivity to increased data variability and potential overfitting. InceptionV3 consistently records the lowest performance in both settings, indicating limited suitability for the dataset. Overall, the table highlights the effectiveness of augmentation for enhancing model robustness while emphasizing EfficientNetB0, followed by ResNet50 and DenseNet121, as the most reliable models for sugarcane leaf disease classification.

Table 4. Comparative Analysis of Baseline Models

Model	Augmentation	Accuracy	Precision	Recall	F1-Score
VGG16 (Simonyan & Zisserman, 2014)	No	0.78	0.78	0.78	0.78
	Yes	0.76	0.76	0.75	0.75
VGG19 (Simonyan & Zisserman, 2014)	No	0.78	0.78	0.78	0.78
	Yes	0.76	0.76	0.76	0.76
ResNet50 (He et al., 2016)	No	0.84	0.84	0.84	0.84
	Yes	0.85	0.85	0.85	0.85
InceptionV3 (Szegedy et al., 2016)	No	0.58	0.58	0.58	0.57
	Yes	0.59	0.59	0.58	0.58
MobileNetV2 (Sandler et al., 2018)	No	0.65	0.65	0.65	0.65
	Yes	0.73	0.73	0.73	0.73
DenseNet121 (Huang et al., 2017)	No	0.65	0.65	0.65	0.65
	Yes	0.83	0.83	0.83	0.83
EfficientNetB0 (Tan & Le, 2019)	No	0.84	0.84	0.84	0.84
	Yes	0.89	0.89	0.89	0.89

Table 5 presents a computational analysis of different pre-trained models in terms of parameter size, inference time per image, and overall model size. The results highlight that lightweight architectures such as MobileNetV2 and EfficientNetB0 offer better computational efficiency, while VGG models are significantly heavier and resource intensive. To validate whether observed performance differences between models are statistically significant, each experiment was repeated five times with different random initialization seeds. Mean accuracy and standard deviation were computed. Additionally, paired t-tests were conducted between EfficientNetB0 and other architectures at a 95% confidence level. Results confirm that EfficientNetB0 significantly outperforms other models ($p < 0.05$), validating the robustness of the observed improvements.

Table 5. Computational Analysis with Baseline models

Model	Parameters (M)	Inference Time per Image (ms)	Model Size (MB)
VGG16 (Simonyan & Zisserman, 2014)	138	15–20 ms	528 MB
VGG19 (Simonyan & Zisserman, 2014)	144	18–22 ms	549 MB
ResNet50 (He et al., 2016)	25.6	8–12 ms	98 MB
InceptionV3 (Szegedy et al., 2016)	23.8	10–15 ms	92 MB
MobileNetV2 (Sandler et al., 2018)	3.4	3–6 ms	14 MB
DenseNet121 (Huang et al., 2017)	8.0	7–10 ms	33 MB
EfficientNetB0 (Tan & Le, 2019)	5.3	5–8 ms	20 MB

4. Discussion

The experimental results demonstrate notable variability in the performance of pre-trained deep learning models when applied to sugarcane leaf disease classification, emphasizing the importance of model selection and training configuration. Although the dataset represents real-field conditions, its moderate size may limit the generalizability of findings across diverse geographical regions and seasonal variations. The 95% confidence interval for EfficientNetB0 accuracy was 0.89 ± 0.012, indicating low variance and strong stability across repeated trials. Additional experiments was done but higher learning rates (0.01) cause unstable convergence in VGG and Inception models, supporting the selection of 0.001 as an optimal compromise between convergence speed and stability. Among all evaluated architectures, EfficientNetB0 consistently achieved superior performance across accuracy, precision, recall, and F1-score metrics. Its compound scaling strategy enabled effective feature learning while maintaining a balanced model capacity, resulting in strong generalization even with a moderately sized dataset. ResNet50 and DenseNet121 also exhibited robust and stable performance, indicating their effectiveness in capturing complex disease-related features such as lesion textures, discoloration patterns, and vein distortions. These models benefited from architectural mechanisms such as residual and dense connections, which facilitated efficient gradient propagation and feature reuse, leading to reduced overfitting compared to deeper, parameter-heavy networks like VGG16 and VGG19.

The analysis further revealed that data augmentation played a critical role in improving generalization across most models. Augmented training reduced overfitting, stabilized validation performance, and improved convergence behavior, particularly for models with higher parameter counts. However, increasing the number of training epochs beyond 50 did not yield further performance improvements; in fact, certain models such as VGG16, VGG19, and MobileNetV2 exhibited slight performance degradation due to overfitting. Additionally, variations in learning rates showed minimal impact on performance, with 0.001 consistently emerging as the most effective setting. These observations highlight that optimal architectural design and balanced capacity contribute more significantly to classification performance than aggressive hyperparameter tuning. In contrast, InceptionV3 demonstrated limited suitability for the given dataset, likely due to its architectural complexity and dependency on large-scale data for effective feature learning.

Despite the promising outcomes, this study has several limitations that warrant discussion. First, the dataset size, although representative of real-field conditions, remains relatively moderate and may not fully capture the wide variability of sugarcane disease manifestations across different regions, seasons, and environmental conditions. Second, the study focused exclusively on image-based classification and did not incorporate additional contextual information such as climatic, soil, or temporal data, which could enhance disease diagnosis accuracy. Third, only standard pre-trained architectures were evaluated, and the models were not explicitly customized to capture sugarcane-specific disease characteristics at a structural level. Finally, computational constraints limit the exploration of more advanced architecture such as large-scale Vision Transformers or extensive ensemble frameworks. Addressing these limitations in future work through dataset expansion, multi-modal learning, and the development of crop-specific deep learning models can further improve robustness, scalability, and real-world applicability of automated sugarcane leaf disease classification systems.

5. Conclusions

This research presented a comprehensive variability analysis of widely used pre-trained deep learning models for sugarcane leaf disease classification under uniform experimental conditions. The comparative evaluation demonstrated that EfficientNetB0 consistently outperformed all other models, achieving up to 0.89 accuracy, precision, recall, and F1-score at 50 training epochs with an optimal learning rate of 0.001, highlighting its superior feature representation capability and parameter efficiency. ResNet50 and DenseNet121 also exhibited strong and stable performance, with classification accuracies ranging between 0.82 and 0.85, making them reliable alternatives for practical agricultural deployment. In contrast, InceptionV3 yielded the lowest performance, with accuracy values between 0.54 and 0.59, indicating limited suitability for the given dataset. Furthermore, increasing the number of training epochs beyond 50 did not result in performance gains; instead, certain models such as VGG16, VGG19, and MobileNetV2 experienced marginal accuracy degradation due to overfitting. Variations in learning rates (0.01 and

0.0001) produced negligible improvements compared to the optimal setting of 0.001, reinforcing the importance of balanced hyperparameter selection. The findings suggest that EfficientNetB0 can serve as a practical backbone for automated sugarcane disease monitoring systems deployed in precision agriculture frameworks.

EfficientNetB0 and MobileNetV2 demonstrate suitability for deployment on edge devices such as NVIDIA Jetson Nano and mobile-based agricultural diagnostic systems due to their optimized parameter efficiency and manageable inference latency. Future work will focus on (1) developing a sugarcane-specific lightweight architecture, (2) incorporating attention mechanisms for lesion localization, (3) performing k-fold cross-validation to improve statistical reliability, and (4) integrating the model into a real-time mobile-based disease advisory system.

Funding: This research received no external funding.

Data Availability Statement: In this section, please provide details regarding where data supporting reported results can be found, including links to publicly archived datasets analyzed or generated during the study. You might choose to exclude this statement if the study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. M. S. Islam et al., "Detection of sugarcane leaf diseases using custom CNN and ResNet50," Lecture Notes in Networks and Systems, vol. 1410, pp. 277–287, 2025, doi: 10.1007/978-981-96-6303-3_20.

2. D. J. Richter, M. I. Bappi, S. S. Kolekar, and K. Kim, "A systematic review of the current state of transfer learning accelerated CNN-based plant leaf disease classification," IEEE Access, vol. 13, pp. 116262–116303, 2025, doi: 10.1109/ACCESS.2025.3584404.

3. D. J. Richter and K. Kim, "Assessing the performance of domain-specific models for plant leaf disease classification: A comprehensive benchmark of transfer learning on open datasets," Scientific Reports, vol. 15, no. 1, pp. 1–31, 2025, doi: 10.1038/s41598-025-03235-w.

4. C. Sun, J. Dou, B. He, Y. Cai, and C. Zou, "WT-ResNet: A non-destructive method for determining nitrogen, phosphorus, and potassium content of sugarcane leaves based on leaf images," Agriculture, vol. 15, no. 16, pp. 1–18, 2025, doi: 10.3390/agriculture15161752.

5. A. Kumar and G. Saini, "A comparative study of deep learning approaches for early detection of sugarcane diseases," Procedia Computer Science, vol. 260, pp. 182–190, 2025, doi: 10.1016/j.procs.2025.03.192.

6. T. Angamuthu and A. S. Arunachalam, "Comparative analysis of deep learning and optimization techniques for sugarcane disease classification," Premier Journal of Computer Science, pp. 1–8, 2025, doi: 10.70389/pjcs.100011.

7. M. I. Hossen, M. Awrangjeb, S. Pan, and A. Al Mamun, "Transfer learning in agriculture: A review," Artificial Intelligence Review, vol. 58, no. 4, 2025, doi: 10.1007/s10462-024-11081-x.

8. S. Srinivasan et al., "Sugarcane leaf disease classification using deep neural network approach," BMC Plant Biology, vol. 25, no. 1, 2025, doi: 10.1186/s12870-025-06289-0.

9. M. A. Mohammed et al., "Edge–cloud remote sensing data-based plant disease detection using deep neural networks with transfer learning," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 11219–11229, 2024, doi: 10.1109/JSTARS.2024.3410515.

10. J. Mahale, S. Degadwala and D. Vyas, "Crop Prediction System based on Soil and Weather Characteristics," 2022 Sixth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Dharan, Nepal, 2022, pp. 340-345, doi: 10.1109/I-SMAC55078.2022.9987366.

11. F. Patel, S. Mewada, S. Degadwala and D. Vyas, "Exploring Transfer Learning Models for Multi-Class Classification of Infected Date Palm Leaves," 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 307-312, doi: 10.1109/ICSSAS57918.2023.10331746.

12. S. Degadwala, D. Vyas, S. Panesar, D. Ebenezer, D. D. Pandya and V. D. Shah, "Revolutionizing Hops Plant Disease Classification: Harnessing the Power of Transfer Learning," 2023 International Conference on Sustainable Communication Networks and Application (ICSCNA), Theni, India, 2023, pp. 1706-1711, doi: 10.1109/ICSCNA58489.2023.10370692.

13. İ. Kunduracıoğlu and İ. Paçal, "Deep learning-based disease detection in sugarcane leaves: Evaluating EfficientNet models," Journal of Operations Intelligence, vol. 2, no. 1, pp. 321–235, 2024, doi: 10.31181/jopi21202423.

14. K. Hu, H. Li, X. Fu, and S. Zhou, "Customized weighted ensemble of modified transfer learning models for the detection of sugarcane leaf diseases," Electronics, vol. 13, no. 23, 2024, doi: 10.3390/electronics13234715.

15. S. D. Daphal and S. M. Koli, "Enhanced deep learning technique for sugarcane leaf disease classification and mobile application integration," Heliyon, vol. 10, no. 8, p. e29438, 2024, doi: 10.1016/j.heliyon.2024.e29438.

16. M. P. Patil, "Deep learning approaches for the detection, classification, and analysis of sugarcane leaf disease," Foundry Journal, vol. 27, no. 5, pp. 18–28, 2024.

17. E. K. Waters et al., "Sugarcane health monitoring with satellite spectroscopy and machine learning: A review," arXiv preprint arXiv:2404.16844, 2024, doi: 10.48550/arXiv.2404.16844.

18. S. Thite et al., "Sugarcane leaf dataset: A dataset for disease detection and classification for machine learning applications," Data in Brief, vol. 53, p. 110268, 2024, doi: 10.1016/j.dib.2024.110268.

19. W. B. Demilie, "Plant disease detection and classification techniques: A comparative study of performances," Journal of Big Data, vol. 11, no. 1, 2024, doi: 10.1186/s40537-023-00863-9.

20. R. Kursun et al., "Classification of sugarcane leaf disease with AlexNet model," in Proc. Int. Conf. Intelligent Systems and New Applications, 2024, pp. 32–37, doi: 10.58190/icisna.2024.86.

21. R. P. Ethiraj and K. Paranjothi, "A deep learning-based approach for early detection of disease in sugarcane plants: An explainable artificial intelligence model," IAES Int. J. Artificial Intelligence, vol. 13, no. 1, pp. 974–983, 2024, doi: 10.11591/ijai.v13.i1.pp974-983.

22. X. Li et al., "SLViT: Shuffle-convolution-based lightweight vision transformer for effective diagnosis of sugarcane leaf diseases," Journal of King Saud University – Computer and Information Sciences, vol. 35, no. 6, p. 101401, 2023, doi: 10.1016/j.jksuci.2022.09.013.

23. C. Sun et al., "SE-VisionTransformer: Hybrid network for diagnosing sugarcane leaf diseases based on attention mechanism," Sensors, vol. 23, no. 20, 2023, doi: 10.3390/s23208529.

24. Y. Huang et al., "Evaluating data augmentation effects on the recognition of sugarcane leaf spot," Agriculture, vol. 12, no. 12, p. 1997, 2022, doi: 10.3390/agriculture12121997.