Hybrid DenseNet–ViT Architecture for Efficient Sugarcane Leaf Disease Detection

Submitted By

MD Aminul Islam Rasel

ID: 221-35-829

Department of Software Engineering

Daffodil International University

Supervised by

Md Hafizul Imran

Assistant Professor

Department of Software Engineering

Daffodil International University

A thesis submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Software Engineering

Fall-2025

Hybrid DenseNet–ViT Architecture for Efficient Sugarcane Leaf Disease Detection

MD Aminul Islam Rasel

Bachelor of Science

DAFFODIL INTERNATIONAL UNIVERSITY

DAFFODIL INTERNATIONAL UNIVERSITY

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	MD Aminul Islam Rasel
Date of Birth	:	03/02/2002
Title	:	Hybrid DenseNet–ViT Architecture for Efficient Sugarcane Leaf Disease Detection
Academic Session	:	Fall 2025

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997) *
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done) *
☐	OPEN ACCESS	I agree that my thesis to be published as online open access (Full Text)

I acknowledge that Daffodil International University reserves the following rights:

1. The Thesis is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the thesis for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the thesis for academic exchange.

Certified by:

(Student's Signature)

Student ID: 221-35-829
Date:13/12/2025

(Supervisor's Signature)

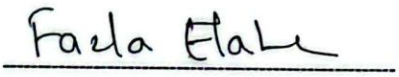
Name of Supervisor
Date: 13/12/2025

NOTE: * If the thesis is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

APPROVAL

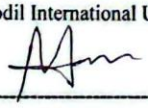
This thesis titled on "**Hybrid DenseNet–ViT Architecture for Efficient Sugarcane Leaf Disease Detection**", submitted by **MD Aminul Islam Rasel (ID: 221-35-829)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS



Chairman

Dr. Fazla Ealhe
Assistant Professor & Associate Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

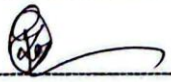


Internal Examiner 1

Dr. Marzia Ahmed
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

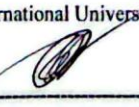
Internal Examiner 2

Dr. Shabnom Mustary
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University



Internal Examiner 3

Md. Rajib Mia
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University



External Examiner

Mohammod Abul Kashem, PhD
Professor
Department of Computer Science and Engineering
DUET, Bangladesh

THESIS DECLARATION LETTER (OPTIONAL)

Librarian,
Daffodil International University,
Daffodil Smart City,
Ashulia.Dhaka,Bangladesh

Dear Sir,

CLASSIFICATION OF THESIS AS RESTRICTED

Please be informed that the following thesis is classified as RESTRICTED for a period of three (3) years from the date of this letter. The reasons for this classification are as listed below.

Author's Name
Thesis Title

Reasons (i)

(ii)

(iii)

Thank you.

Yours faithfully,

_____
(Supervisor's Signature)

Date:

Stamp:

Note: This letter should be written by the supervisor and addressed to the Librarian, *Daffodil International University* with its copy attached to the thesis.

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and, in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.

(Supervisor's Signature)

Full Name : Md Hafizul Imran

Position : Assistant Professor

Date : 27/11/2025

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.

_____

(Student's Signature)

Full Name : MD Aminul Islam Rasel

ID Number : 221-35-829

Date : 27 November, 2025

Hybrid DenseNet–ViT Architecture for Efficient Sugarcane Leaf Disease Detection

MD Aminul Islam Rasel

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Science/Master of Science

Department of Software Engineering (Major in Data Science)

DAFFODIL INTERNATIONAL UNIVERSITY

November, 2025

ACKNOWLEDGEMENTS

The authors would like to extend their sincere thanks to the Software Engineering Department of Daffodil International University, Dhaka, Bangladesh for providing their full support and resource without which this research could not have been accomplished. We are grateful to **Md Hafizul Imran** for his useful suggestions, comments, and continued support during the whole research work. His guidance has been indispensable to this work being performed well.

We are also thankful to the Kaggle community for sharing its publicly available datasets, used as dataset for experimental analysis provided in this article. The dataset has been very helpful in our work and played an important role toward designing the proposed approach. Thanks to the colleagues and reviewers from various workshops for their meaningful discussions and constructive feedback that helped improve significantly on quality and clarity of this hybrid DenseNet–ViT architecture. Their knowledge and critical insights have been invaluable in helping us to improve our work. Thank you to our families for all of the emotional and logistical support. For their comprehension, support and patience during my research period to make it possible for us to fully commit ourselves in this project. This study would not have happened without their unflappable faith in us.

Finally, all the errors and omissions are entirely the authors own. We are dedicated to further development and learning from our experiences with this work.

DEDICATION

This work is for my family, who have supported me every step of the way, and loved and encouraged me through thick and thin. Their patience, comprehension and self-sacrifice during this process are of absolute importance to the realisation of this research. The faith that they have in me is the basis of this work.

I also dedicate this thesis to all the sweaty farmers in the world, and especially those growing sugarcane. Their day-to-day work, their tenacity and dedication to feeding the world are a never-ending source of inspiration for the innovations challenging to change agricultural practices, boost yields and guarantee that our food supply is secure. Their hard work is the glue that holds our society together and this work celebrates their immeasurable accomplishments. Last but not least, I dedicate this work to all who are striving to create a better world with both technology and sustainable solutions. In particular, I dedicate it to my fellow members in the field of precision agriculture which is changing the way we think about farming with technology. Based on these thoughtful visions, it won't be long before these breakthroughs are able to unlock sustained, positive impact on communities and the environment that will sustain our future generations.

ABSTRACT

Sugarcane is an important tropical agricultural crop, which is under the threat of several leaf diseases like mosaic disease, red rot, rust, yellow leaf and mawa (banded blight) that causes a huge loss in yield. Traditional manual inspection-based disease detection is time-consuming and error-prone, it cannot be used for many varieties of large-scale monitoring. This paper demonstrates a new approach based on deep learning to automatically identify six sugarcane leaf diseases. We experimented with several deep learning-based model, such as EfficientNet-B0, DenseNet-121, MobileNetV2, ResNet-50 and hybrid architecture which is a fusion of dense net and resnet and vision transformer (ViT). Results revealed that the accuracy with CNN-based models worked well in >90%. The performance result of the Hybrid super model between DenseNet and ViT is also significantly better than others with 96.7% validation accuracy, 96.5% test accuracy and F1 score weighted=0.97(Far from them). This shows the potentiality to integrate local convolutional features with global context features from transformers for enhancing classification performance. This type of a solution could offer an efficient scalability and reliability for precision agriculture and contribute to reducing reliance on manual disease detection and to sustainable crop growing. Moreover, the current method is suitable for detecting in real-time and practical for farmers to eliminate disease at early stage of crop plant filed, so it can greatly limit area-affected plants and agricultural production. Incorporating deep learning into disease management is a part of the global trend toward “smart” farming, which intends to use technology-enabled tools for more precise, information-based decision-making. 2 Significance statement This paper opens up the road towards revolutionizing crop management and ensure food security in tropical zones by leveraging AI-based solutions.

Keywords: Deep Learning; Vision Transformer; DenseNet; Sugarcane Leaf Disease Detection; Precision Agriculture; Explainable AI

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATION

CHAPTER 1

INTRODUCTION

1.1.1 OVERVIEW

Sugarcane (*Saccharum officinarum*) continues to be an important tropical agriculture commodity world- wide which is grown in more than 100 countries at a production of over 1.9 billion tons per year, and has essential supply chains for sugar, ethanol and bioenergy generations [1] [2]. It not only sustains livelihoods for the millions of subsistence smallholder farmers in regions like India, Brazil and Southeast Asia, but it offers crucial energy promotion security. Its broad applications ranging from food to biofuels categorize it as an essential crop in meeting the challenges of feeding the population and providing renewable energy. Yet, it is severely threatened by numerous diseases of which the majority inflict serious and consistent damage to productivity. When not controlled, these pathogens have strong negative impact on plant growth and sucrose accumulation leading to yield losses fluctuating between 60-80% in severe epidemics [3] [4].Among these pathogens, red rot caused by *Colletotrichum falcatum* is devastating and often leaves the whole field uncultivated and cost serious loss of high yielding cultivars from cultivation [5]. Rust, mosaic, mawa and yellow leaf syndrome are other important diseases which exhibit same or mis-leading symptoms. For instance the early stage mosaic symptoms can be seen as faint chlorotic streaks that are similar to nutritional deficiency and rust pustules can create additional complexity for real-time field level identification [6] [7]. Such overlapping symptoms in- crease the likelihood of failure to recognize and handle timely such diseases, leading to the spread of disease till the infection has advanced beyond its optimum stage for intervention and that economic loss is unavoidable.

1.1.2 BACKGROUND

Conventional surveillance practices that depend on roving eye scouting and laboratory identification are laborious, error-prone and rarely practical for big area's or geographically spread plantations [8]. In extremely isolated and rural farming environments, availability of more sophisticated diagnostics, including PCR or

ELISA is limited, and thus difficult to perform an early diagnosis of the disease. Even experts struggle to find infections that live deep within dense foliage. It is this late diagnosis besides early detection of the diseases constitutes an urge to researchers for developing vision based automatic systems deploying machine learning (ML) and deep learning (DL) techniques to predict and diagnose the disease in real-time [9] [10].

On the other hand, traditional ML methods might rely on engineered features which are susceptible to real-world variability (e.g., lighting conditions, background noise or weather) and prone to being broken in adversarial set- tings in comparison current DL approaches such as CNN provide a more `solid' mechanism. For example, leaf images can be used to harvest the hierarchical structures of plant tissues by the pattern information for each layer on CNN- architectures (e.g., Densenet or Resnet), which is robust to variations in background environments [11]. But this advance creates a new difficulty: how do we get these systems to work in all kinds of real-world situations without one or the other trading off accuracy? Even more, recent advances well-supersede common RGB image acquisition. Hyperspectral imaging spanning a wide range of wavelengths can be valuable in revealing subtle biochemical responses of plants that are reflected as modifications in pigmentation and the moisture content (an early onset disease symptom). These interventions might be able to recognize disease before it becomes particularly severe and prevent serious outbreaks. However, the high cost of a hyperspectral image system has resulted in constrained application scenarios and failed to satisfy the smallholder farmer's need in resource-limited regions [12].

1.1.3 PROBLEM STATEMENT

Traditional plant health monitoring still depends largely on field scouting and laboratory testing. These techniques are tedious, time-consuming and not suitable for large or widespread plantations [8]. In isolated rural areas with farming for a main occupation, diagnostic methods such as PCR and ELISA are not accessible and timely diagnosis is particularly difficult. Even experienced agronomists have difficulty spotting early infections buried under thick greenery. This ongoing failure of diagnosis constitutes a major hurdle towards combatting will malady and minimizing crop damage.

In this study, we proposed a novel dense block-based deep learning model (DenseNet-121–ViT) for the prioritized multi-classification of sugarcane leaf diseases and provided its effective end-to-end application that can perform reliably under harsh and diverse field conditions. For example, DenseNet-121 might have the capacity to capture fine-grained local responses that would provide clue for detecting the beginning of alternance in a plant such as saturated rust spots or mosaics. On the other hand, ViT's self-attention mechanism can aggregate global information to detect relatively obscure symptoms that are overlooked by local features. The suggested system identifies six sugarcane leaf diseases Healthy, Mawa, Mosaic, Red-Rot, Rust and Yellow-leaf.

1.1.4 PROPOSED METHOD

This paper presents a hybrid DenseNet-121–Vision Transformer (ViT) model specifically developed to perform multi-class sugarcane disease classification in various real-world settings. First, DenseNet-121 is responsible to collect fine-grained local features such as rust pustules or mosaic streaks and ViT provides a global grasp of the leaf structure by self-attention. The model provides classification of six classes: Healthy, Mawa, Mosaic, Red Rot, Rust and Yellow Leaf. For interpretability, Grad-CAM heatmaps illustrate the areas on a leaf that are affecting each prediction. These visualizations validate model decisions and are training materials for agricultural experts. In experiments on an in-field dataset, the hybrid model achieved **96.7% validation accuracy** with an inference time of under 80 ms, enabling real-time diagnostics on consumer-grade web app devices.

1.1.5 MOTIVATION

The combination of symptom similarity, late diagnosis and the unavailability of advanced laboratory tools highlights the need for a dependable, low-cost and field-ready disease detection system. Farmers and extension workers require a method that performs reliably in varying backgrounds, inconsistent light and real cultivation environments.

1.1.6 CONTRIBUTIONS

This study makes several contributions toward improving automated sugarcane disease diagnosis. It introduces a hybrid DenseNet121–ViT architecture that brings together convolutional feature extraction and transformer-based global attention, enabling more reliable classification across six major sugarcane leaf diseases. A dataset of over 2,800 labelled images was prepared and standardized through consistent preprocessing and augmentation to reflect real field conditions. The model's performance was examined through a wide set of evaluation metrics, including accuracy, precision, recall, F1-score, and ROC–AUC, demonstrating clear improvements over individual DenseNet and ViT baselines. The computational aspects of the system were also analysed, covering runtime behaviour, GPU memory usage, model size, and inference speed. To support interpretability, Grad-CAM visualizations were generated to highlight the image regions most relevant to each prediction. Finally, the complete system was deployed using a Flask-based web interface, allowing real-time image uploads and predictions, which shows the practical usability of the proposed approach in field-ready decision-support tools.

1.1.7 RESEARCH OBJECTIVE:

1. To propose a powerful hybrid DenseNet-121–ViT model for precise sugarcane disease classification by taking into consideration field variation and real-world constraints.
2. To assess model performance such as precision, recall, F1-score and AUC metric to guarantee the efficiency of disease classification in all the six categories.
3. To include visual explanations for model predictions via Grad-CAM, allowing interpretability and decision-making support for stakeholders.
4. To facilitate scalable and affordable deployment of the model with mobile apps, drone-based systems supporting sustainable agronomy for smallholder farmer.

CHAPTER 2

LITERATURE REVIEW

2.1.1 TRADITIONAL MACHINE LEARNING APPROACHES

The early works in sugar cane disease recognition were mostly based on common machine learning approaches that relied on hand-crafted features from color, multispectral, or hyperspectral images. A comprehensive review [13]studied different classifiers such as SVM, KNN, and random forests on red rot and smut. Most of these methods worked well in a lab environment. Still, they were known to be less effective in real-world agricultural environments, as lighting and background noise vary from the normal distribution setting. This unstable class distribution can occur, such as how machines are maintained or used by the farmer, etc. (the fragility of hand-crafted feature design considering these dynamics). On top of this work, [7]designed a smartphone-compatible pipeline utilizing an HMM with anisotropic diffusion for filtering, specifically focusing on rust, yellow spot, and ring spot. This pragmatic approach has produced valuable results in the field. Whenever phone images are of low quality, the emphasis is still placed upon usability in the field. Alternatively, [14]developed a knowledge-based system through CLIPS and Delphi for rule-based diagnosis of mosaic or red rot based on visual symptoms. Despite its perceived common sense for non-specialists, its hard- wired rules lacked adaptation to rarer disease presentations and therefore fostered an interest in more flexible learning-based approaches. Early attempts also focused on simple pipelines of image analysis (e.g., one in [6]that used segmentation followed by classification), but were limited by parameter dependencies. These same design–adaptability trade-offs are evident in analogous engineering-driven systems beyond agriculture. For instance, in the field of robotics, [15]introduced PIRATE, a pipe inspection robot, which illustrates the need to integrate mechanical precision with adaptive algorithms for robustness in uncertain environments. This focus on robustness and field- ready user satisfaction reflects limitations experienced by early rule-based

sugarcane disease systems, for example, as noted in [14], where environmental variability typically led to performance deterioration.

2.1.2 DEEP LEARNING INNOVATIONS

Moreover, when the shift to deep learning, in particular CNNs, became dominant in the field, taking into account hierarchical features directly from input image data without manual engineering was revolutionary. [1]developed CNN architectures on 20K sugarcane leaf images to differentiate mosaic from red rot and have shown that CNN models can also capture intricate patterns, but reminded the requirement of 'trust by end-users like the farmers' considering low-to-zero interpretability. proposed a Diversified Deep Learning Architecture (DDLA) to address overlapping symptoms, using as inputs discrete wavelet transform (DWT) and histogram equalization, with results of 97% accuracy on training data and 87% on unseen test under mosaic subtypes—calling for a faster alternative than traditional laboratory PCR tests. To detect red stripes within seconds through a custom-designed YOLO framework [5], for rapid processing and adapt- able deployment across different relational hardware devices. More advanced base models were also more popular, as [5]tested 6,748 images across 11 categories and found that EfficientNet-B6 performed best. Similarly, in validation experiments, the top contributors have been EfficientNet-B7 and DenseNet201 [11], but it has been cautioned that deeper networks do not always lead to better performance in low-resource settings. Weighted ensembles enhanced robustness; aggregated transfer learnt models for binary tasks to enhance F1-scores while reducing computational cost in the presence of multiple labels. Interpretability can be found in, [16]which introduced a DenseNet-SVM model with LIME visualization (which not only delivered a diagnosis but also advice). On the contrary, Chavan et al. [17]showed that a simple sequential CNN could outperform DenseNet (94% to 75%) on a dataset with a smaller number of samples, and promote that simpler architectures are preferable when data is scarce. Further- more, in the Philippine setting, no study has used digital methods and machine learning to classify local microalgae and cyanobacteria species. This is the first national project that uses OpenCV preprocessing and a CNN structured by TensorFlow to create a recognition tool based on algae widespread in freshwater [18]. Relevant applications where features were extracted from gray-level co-occurrence matrix (GLCM) with k-means clustering to estimate

cutworm damage severity on leaves of peachy were as follows: 88.31% accuracy [19]. Beyond agriculture, artificial neural networks (ANNs) have automated wood type recognition from macro photos, achieving an F1-score of 87.9% [18] and the U-Net has located giant epithelial cells in measles cytology slides [19]These examples reflect the versatile applicability of CNNs and ANNs in bio-agricultural-related work, which motivated the hybrid pipeline proposed in this work. The optimization method was essential: [20]demonstrated the Enhanced Environmental Adaptation Method (EEAM) to fine-tune CNNs, outperforming genetic algorithm baselines. [21] [22]combined CNN predictions with GLCM-based texture features, for seven pathologies to balance accuracy and inference time, which is essential for on-the-field deployment of the system. Stacked architectures such as perceptron segmentation reported in [22] segmentation sharpened multi-class predictions. The need to make the real world deployable has also been emphasized in other fields. E.g., [23] presented NISHASH, an inexpensive mechanical ventilator which was developed in the COVID-19 crisis to show how deep learning and embedded control can allow automation even on a budget. Such design heuristics guide the development of field deployable CNN architectures for resource-limited agricultural take up.

2.1.3 ADVANCED IMAGING AND HYBRID APPROACHES

Early detection is possible via hyperspectral imaging of spectral signatures undetectable by the unaided eye. They used such data to spot smut before symptoms, which came as minimal shifts in spectral features extracted from a suitably dual attention ResNet model [12]. In the context of ad- dressing the class imbalance [24], for the labelling, stacked InceptionV3, SqueezeNet and Deep learning models were utilized whereby it [25] [11]combined pretrained CNN models (InceptionV3 and VGG) with SVM classifiers, yielding high AUC values. Miniaturization of field portable tools was improved by [8]using probabilistic low memory algorithms. Beyond imaging, [4]linked deep diagnostics with biological controls for red rot and [2] reviewed serological, molecular, and breeding based approaches. Large scale re- mote sensing monitoring was studied in somard2021remote, advocating for the multispectral fusion at drones or satellites, as it is already successful for UAV in citrus disease mapping [26]. Parallel hybridization in the Philippines: OpenCV denoising and classification freshwater algae at 93.75% accuracy

using CNNs (Abriol-Santos, Automated Classification and Identification System for Freshwater Algae Using Convolutional Neural Networks, 2023), and leaf peachy damage analysis with GLCM plus k-means [27], Classification and Percent Severity of Peachy Damage Caused by Cutworm. In other cases, ANN-based image analysis was used to detect local wood species mendoza2022, and U-Net segmented measles-associated cells within smears [28]; suggesting scope for flexible multi-input diagnostic tools. Genetic analyses, including genome-wide association for resistance [29], add additional depth to detection with breeding relevance.

2.1.4 PERFORMANCE METRICS AND CHALLENGES

Accuracy, precision, recall, F1-score and AUC were popular evaluation measures with good performing results from EfficientNet-B7 [11] CNNGLCM integrations [21] and DDLA [30]Common challenges include the scarcity of labeled examples for rare co-occurrences, the danger of overfitting and lower effectiveness in uncontrolled scenarios [8]On-device execution requires compression for limited devices [3] [7].

2.1.5 RESEARCH GAP

1. Dataset Constraints: Paucity of diverse annotated samples for rare or mixed infections across the population hinders its extensive usage [2] [13].

2. Multi-Modal Integration: The integration of hyperspectral data for the soil with context such as weather or soil, through IoT is an underexplored area [12] [8].

3. Explainability Gaps: There will be need for interpretability tools more advanced than LIME to win users' trust [10].

4. Real-world validation: Limited high-scale field testing in low-resource areas [4] [14]

5.Climate Resilience: Updated model for new strains in the face of environmental variations [2] [30]

6.Edge Efficiency: Sparse benchmarks on methods such as EEAM with ensembles on edge devices [20] [21]

2.1.6 COMPARISON BETWEEN EXISTING WORKS

Table 1: Comprehensive Review of Sugarcane Disease Detection Methods

Ref.	Main Objective	Methodology	Dataset	Key Findings/Accuracy
[16]	XAI for early detection	DenseNet-SVM with LIME	Not specified	Explainable warnings with pesticide advice
[9]	Novel DL framework	InceptionV3, VGG-16/19, SVM	Not specified	90.2% AUC (VGG-16+SVM)
[13]	ML techniques review	SVM, KNN, Random Forests	RGB/multi/hyper	Real world limitations identified
[5]	EfficientNet evaluation	EfficientNetV1/V2 variants	6,748 images (11classes)	Outperforms baseline CNNs
[8]	Mobile detection systems	ML/DL hybrids	Diverse plant sets	Automation efficiency issues
[3]	Red stripe via YOLO	YOLO CNN	4,000images (balanced)	2s inference, high precision
[7]	leaf spot system	HMM + anisotropic diffusion	Mobile captures	75% post-preprocessing
[11]	DNN classification	EfficientNet, DenseNet, etc.	6,748images(SLD)	EfficientNet-B7/DenseNet201 top
[17]	DL comparison	DenseNet vs. Sequential	Limited custom	Sequential at 94%
[6]	Image processing	Segmentation + classification	Custom images	Threshold-based identification
[26]	UAV yield/disease es- limitation	CNN on multispectral	CitrusUAV data	High accuracy, transferable

Table 1 presents the details of different studies conducted on sugarcane disease detection based on ML/DL techniques. Various algorithms, including DenseNet-SVM34, YOLO CNN35 and EfficientNet,36 have been applied to enhance the accuracy and efficiency. There is research on interpretable AI, gap detection systems, hybrid model as well as ingestion to RGB, multispectral or self-generated images. The results are mixed, with models like the EfficientNet-B7 and DenseNet201 setting new state-of-the-art. Overall, such studies aid in the development of tools for disease diagnosis suitable for application in science and precision agriculture.

CHAPTER 3

METHODOLOGY

In this dissertation, a novel deep learning model for sugarcane leaf disease diagnosis was employed that fuses the merits of CNNs and Transformer models. Specifically, we introduce a new hybrid model that integrates DenseNet-121, a popular CNN with high feature reuse and image classification accuracy, and Vision Transformer (ViT), an efficient method for capturing long-range dependencies and global context between the pixels in images. This fusion enhances the model's discriminative power to facilitate an accurate recognition for sugarcane leaf diseases in that the local and global information are fully utilized so as to improve the overall detection performance.

As per the structural methodology, the framework proceeds in several stages. The dataset preparation can mainly be divided into data collection, cleaning, and preprocessing to provide high-quality input to the model. This is a very important detail to build a solid base for the remaining runs. A feature extraction step comes next, using DenseNet-121 model to extract hierarchical and discriminative features of input images. These features are subsequently fed to the ViT for learning global patterns and dependencies, that may be present in-order to recognize complex disease symptoms effectively.

3.1.1 DATASET PREPARATION

The dataset that we used for this study was obtained from a public Kaggle dataset (https://www.kaggle.com/datasets/mdaminulislamrasel/sugarcane-dataset-lasted) and comprised over 2,800 annotated sugarcane leaves images. These images were then categorized into six classes namely Healthy, Mawa, Red Rot, Yellow Leaf, Mosaic and Rust. It therefore contains a large spectrum of symptoms that can be found in sugarcane fields. These are images of a variety of textures, coloring patterns and lesion layouts which reflect the natural variance within field. This diversity is essential for establishing a model that will generalize in various environmental conditions and diseases. A training, validation and testing dataset was formed by splitting these data. Training, validation and

testing sets are composed of 2,093, 448 and 451 images. The split was random, so that each subset had at least a proportional number of samples of each class. This helps in fair comparison of the model for diverse disease types. We resized all images to be 224x224 pixels to be compatible with the input size of pre-trained DenseNet-121 and Vision Transformer (ViT) models, as these showed excellent performance for image classification. This rescale operation has a side benefit in that its adjusted image will be small enough for the model to be able to handle optimally, while still being large enough maintain indispensable characteristics for an accurate classification.

Figure 1: Sample sugarcane leaf images from the dataset across six disease categories.

3.1.2 FEATURE EXTRACTION

To efficiently classify sugarcane leaf diseases, two feature extractors were employed to learn local and global visual representations.

DenseNet-121: We used a pretrained DenseNet-121 model as the convolutional backbone, which was initialized with ImageNet weights. The weights of all convolutional layers are fixed with the exception of the classifier which was replaced by a linear layer as follows:

$$\text{Classifier} = \text{Linear}(\text{num ftrs}, \text{num classes})$$

where num ftrs = 1024 while num classes = 6, these being Healthy, Mawa, Mosaic, RedRot, Rustand Yellow. For training images, standard augmentations were applied (i.e.,

RandomResizedCrop(224), random horizontal flip and normalization using ImageNet statistics) and validation images were of size 256 by 256 then center-cropped to 224. Fine-tuning of the classifier was performed over 25 epochs with Stochastic Gradient Descent (SGD) using a learning rate of (0.001) and momentum (0.9), where the Cross-Entropy Loss is formulated as:

$$L=-i=1\sum Cyilog(y\hat{}i)$$

where C is the number of classes, yi is the true label, and yˆi is the predicted label.The denseNet-121 achieved a maximum validation accuracy of 91.52 % and there was macro average precision, recall and f1-score for all as,and. The confusion matrix showed great results with some minor mix-ups around the visually similar classes such as Healthy and Mosaic. The model had strong robustness for the recognition of local and textural difference, so it was highly suitable for the determination of leaf disease.

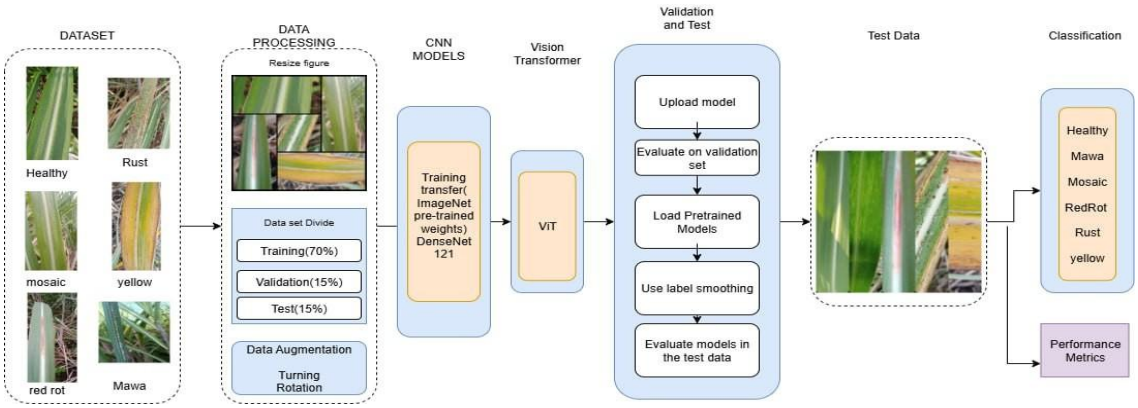


Figure 2: Illustration of the proposed hybrid DenseNet–ViT framework for disease detection.

Vision Transformer: The Vision Transformer model is originally used to learn long-range global dependencies and spatial relations among the image patches. For computational reasons we used a lightweight alternative with ViT pretrained on ImageNet, allowing to remain efficient while keeping generalization. The last classification layer is substituted with:

Equation 1: Classifier = Linear (in features, num classes)

where in features = 1280 and num classes = 6. Data augmentation was: Resize (224), RandomRotation(15°), RandomHorizontalFlip, ColorJitter(brightness/contrast/saturation = 0.3) and RandomResizedCrop(224,

scale= (0.85, 1.0)), whereas validation uses only resized and normalization. The model was fine-tuned using Adam optimizer with learning rate of 1×10^{-5}, weight decay of 1×10^{-4} for 30 epochs with early stopping (patience = 5), loss function used Cross-Entropy Loss:

Equation 2: L=−i=1∑Cyilog(y^i)

and accuracy as:

Equation 3: Acc=Total Samples/∑i=1N1(y^i=yi)

The ViT model obtained the validation accuracy to be **92.19**, while the macro averaged Precision, Recall, F1-score was 0.92. Its confusion matrix revealed extremely consistent classification quality, where the class *Mawa*, instead, performed perfectly in all three measures. This network is good at capturing global structure information and was able to compensate DenseNet's local feature extraction ability, so it already provides a strong baseline for hybrid fusion.

3.1.3 HYBRID MODEL ARCHITECTURE

The proposed hybrid model uses the feature representations from DenseNet-121 and ViT to exploit the complementarity of the benefits of CNN-like and attention-based strategies. Fine-grain local textural details are learned by DenseNet-121, whereas the global spatial and contextual information is provided by ViT. The final output feature maps of the two models were then concatenated together to generate a 1792D joint feature (1024 + 768), followed by a multilayer classifying module for predicting the decision. The classifier architecture is defined as: where num classes = 6 refers to the sugarcane leaf classes; Healthy, Mawa, Mosaic, RedRot, Rust, and Yellow. The loss function was enforced with generalization induced by label smoothing ($\epsilon = 0.1$), which is given by:

Equation 4: L=(1−ϵ)i=1∑Cyilog(y^i)+ϵi=1∑Clog(y^i)

where C = 6 represents the number of categories, yi is the actual label, and y^i indicates the predicted likelihood for class i. Accuracy was computed as:

Equation 5: Acc=total samples∑i=1N1(y^i=yi)×100

Table 2: Hybrid DenseNet121–ViT Model Architecture

Component	Description
DenseNet121 Back-bone	Pretrained CNN, classifier removed (1024-D features)
ViT	Pretrained transformer encoder (768-D features)
Feature Fusion	Concatenation of DenseNet and ViT outputs (1024 + 768 = 1792-D)
Fully Connected Layer	Dense layer with 512 hidden units
Activation	ReLU non-linearity
Dropout	Dropout (rate = 0.4) for regularization
Output Layer	Softmax with 6 neurons (disease classes)

Images were resized to 224 × 224 pixels and normalized with ImageNet mean and standard deviation values before feeding into the network. The dataset was divided into training, validation, and test sets using Image Folder loaders with batch sizes of 32. For model optimization, the Adam algorithm was used with a learning rate of 1 ×10−4, trained for 25 epochs, and Cross-Entropy loss with label smoothing. During the training process, epoch runtime, GPU usage, and validation metrics were recorded moment- by-moment, while accuracy-loss curves were drawn for performance visualization. The proposed hybrid model obtained a macro-avg precision, recall, and F1-score of 0.93, outperforming the single DenseNet121 (0.9152) and ViT (0.9219) models. Analysis of the confusion matrix showed sustained performance and almost no misclassification in all six disease classes. By well fusing the ability to extract local texture features and acquire global contextual relevance, the hybrid framework achieved better discriminability and generalization performance on sugarcane leaf disease recognition.

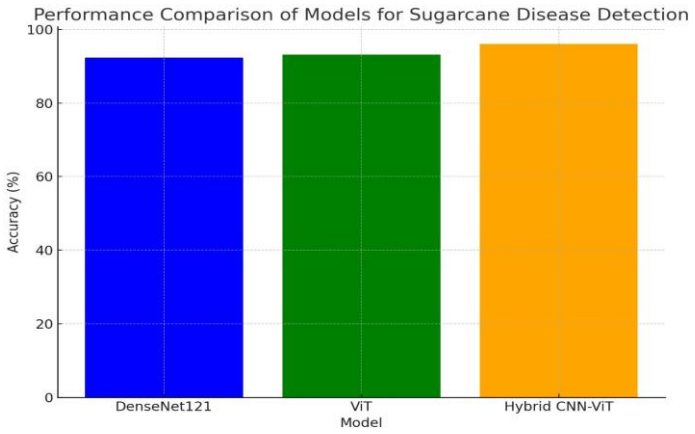


Figure 3: Performance comparison of CNN, transformer, and hybrid models on the sugarcane disease dataset.

We compare three architectures, i.e., DenseNet121, ViT and our proposed hybrid CNN–ViT using the classification accuracy on sugarcane disease data as shown in Fig. 3. Both the CNN and transformer models perform well and are mostly comparable, indicating their capacity of capturing local texture cues and global spatial patterns, respectively. However, the performance improvement of the hybrid model is evident, demonstrating that integrating convolutional feature extraction and transformer global attention achieves a more discriminative representation. This enhancement demonstrates the fusion strategy's capabilities in dealing with rich visual content and subtle variation of symptoms of diseased sugarcane leaves.

3.1.4 ALGORITHM

Algorithm 1: Data Preparation

Require: Image directories (*train/*, *val/*, *test/*); target size: 224 × 224
Output: Mini-batch loaders for training, validation, and testing

1. Define the preprocessing pipeline:
    - Resize images to **224 × 224**
    - Convert to tensor
    - Normalize using ImageNet statistics $\mu = [0.485, 0.456, 0.406]$, $\sigma = [0.229, 0.224, 0.225]$

2. Load the dataset partitions using **ImageFolder**, producing:
$$D_{\text{train}}, D_{\text{val}}, D_{\text{test}}$$

3. Construct data loaders:
    - L_{train}: batch size 32, shuffle enabled
    - L_{val}: batch size 32
    - L_{test}: batch size 32

Algorithm 2: Hybrid Model Forward Pass

(DenseNet-121 and ViT used as frozen feature extractors)
Require:

- Input batch $x \in \mathbb{R}^{B\times 3\times 224\times 224}$
- DenseNet-121 encoder → 1024-D feature vector
- ViT-B/16 pooler output → 768-D embedding
- Number of output classes: C

Output: Class logits $z \in \mathbb{R}^{B\times C}$

1. Freeze backbone computation using **no-grad** mode.
2. Extract convolutional features:
$$f_{\text{DN}} \leftarrow \text{DenseNet}(x)(B \times 1024)$$
3. Extract transformer features (CLS-token embedding):
$$f_{\text{ViT}} \leftarrow \text{ViT}(x)(B \times 768)$$
4. Concatenate feature vectors:
$$f \leftarrow [f_{\text{DN}} \;||\; f_{\text{ViT}}]$$
5. Pass through a fully connected block:
    - Hidden layer:
$$h \leftarrow \text{ReLU}(\text{Linear}(f; 1792 \rightarrow 512))$$
    - Apply dropout ($p = 0.4$)
6. Generate final class logits:
$$z \leftarrow \text{Linear}(h; 512 \rightarrow C)$$
7. **Return** z

Algorithm 3: Train HybridModel with Label Smoothing and Track Best Validation

Require: HybridModel M; optimizer Adam(M, lr=10^{-4}); epochs E=25; label smoothing ϵ=0.1; load- ers L_{train}, L_{val}

Ensure: Best weights $\theta^{\star}$ (by validation accuracy); histories of loss/accuracy, epoch time, peak GPU mem

Initialize histories H_{train}, H_{val}, best acc $\leftarrow 0$, $\theta^{\star} \leftarrow$

for e = 1 to E **do**

Reset GPU peak mem stats; start timer

Train phase: set M to train

$L_{sum} \leftarrow 0$, correct $\leftarrow 0$, total $\leftarrow 0$

for batch (x, y) in L_{train} **do**

$z \leftarrow M(x)$

$\ell \leftarrow$ CrossEntropy(z, y; label smoothing = ϵ)

Backprop: zero grad $\rightarrow \nabla\ell \rightarrow$ step

Accumulate: $L_{sum}\ell \cdot |x|$, correct $\sum$ (arg max z = y), total |x|

Release batch tensors; empty CUDA cache (optional)

end for

TrainLoss $\leftarrow L_{sum}$/total; TrainAcc $\leftarrow 100 \cdot$ correct/total

Append to H_{train}

Validate phase: set M to eval $L_{val} \leftarrow 0$, correct $\leftarrow 0$, total $\leftarrow 0$

Initialize lists Y_{true}, Y_{pred}

for batch (x, y) in L_{val} **do no grad**

$z \leftarrow M(x)$; $\ell \leftarrow$ CrossEntropy(z, y; ϵ)

$L_{val}\ell \cdot |x|$; correct $\sum$ (arg max z = y); total |x|

Append y to Y_{true} and arg max z to Y_{pred}

end for

ValLoss ← L_{val}/total; ValAcc ← 100 · correct/total

Measure epoch time and peak GPU memory; log metrics

if ValAcc > best acc **then**

best acc ← ValAcc; $\theta^{\star}$ ← state(M)

end if

end for

Load $\theta^{\star}$ into M

return $\theta^{\star}$, histories, and last Y_{true}, Y_{pred} (for report/plots)

Algorithm 4: Validation Reports and Curves

Require: Y_{true}, Y_{pred}, class names

Compute classification report (precision/recall/F1 per class; micro/macro/weighted)

Compute confusion matrix and draw heatmap

Plot histories: (i) train/val loss, (ii) train/val accuracy, (iii) peak GPU memory per epoch, (iv) epoch time

Algorithm 5: Evaluate on Test Set

Require: Trained M; loader L_{test}; class names
Ensure: Test classification report, confusion matrix, weighted F1

Set M to eval; initialize Y_{true}, Y_{pred}

for batch (x, y) in L_{test} **do no grad**

z ← M(x); $\hat{y}$ ← arg max z

Append y to Y_{true} and $\hat{y}$ to Y_{pred}

end for

Compute confusion matrix heatmap; classification report (per class); weighted F1

return metrics and plots

Algorithm 6: Single-Image Inference with Confidence Gate

Require: Trained model **M**; input image **I**; preprocessing transforms; confidence threshold $\tau \in (0,1)$
Ensure: Predicted class label or "low-confidence/unknown"

1. Apply preprocessing: $x \leftarrow \text{Transform}(I)$; $x \in 1 \times 3 \times 224 \times 224$.
2. Set **M** to evaluation mode.
3. Compute output: $z = M(x)$; obtain probabilities $p = \text{softmax}(z)$.
4. Determine predicted class and confidence:
   $(c^{\backslash *}, s^{\backslash *}) \leftarrow (\arg\max p, \max p)$.
5. If $s^{\backslash *} > \tau$:
   6. Return the predicted class name and confidence $s^{\backslash *}$.
6. Else:
   8. Return "Low confidence / possibly out-of-distribution" with score $s^{\backslash *}$.
7. End if.

Algorithm 1: Prepares the dataset by resizing images, normalizing, and setting up data loaders for training, validation, and testing with a batch size of 32.Algorithm 2: Describes the forward pass of a hybrid model that integrates DenseNet-121 and Vision Transformer (ViT) for feature extraction, followed by concatenation, activation, dropout, and classification output.In the hybrid model, the DenseNet-121 encoder extracts convolutional features $\mathbf{f}_{DN} \in R^{1024}$, while the ViT-Base encoder produces global attention embeddings $\mathbf{f}_{ViT} \in R^{768}$ from the [CLS] token. Both pre-trained backbones remain frozen to retain their learned representations. These feature vectors are concatenated into a joint embedding $\mathbf{f} = [\mathbf{f}_{DN}; \mathbf{f}_{ViT}] \in R^{1792}$ and passed through a multi-layer classifier defined as:

Equation 6: z =W2(Dropout (ReLU(W1f +b1)))+b2

where $z \in R^C$ represents the output logits for C = 6 sugarcane disease classes. Algorithm 3 (Training and Validation) applies the Adam optimizer with a learning rate $\eta = 10^{-4}$ and employs a smoothed cross-entropy loss:where $\epsilon = 0.1$ is the label-smoothing factor that improves model generalization. Training proceeds for E = 25 epochs, minimizing L and updating model parameters θ. Best parameters $\theta* = \arg$

maxθ Acc val are saved for further consideration and we monitor epoch time, GPU memory consumption, and accuracy. Algorithm 4 (Validation Reporting) measures the performance of model by widely accepted performance metrics such as accuracy, precision, recall and F1-score. The confusion matrix M ∈ RC×C is compiled for visualization of inter-class prediction consistency and training curves are also generated to study convergence and stability.

Algorithm 5 (Test Evaluation) uses the best model on the test set to calculate final weighted F1-score and plot class-level confusion matrices, proving generalization of model. Once a model is optimised based on training and validation data, it is applied to the test set in order to determine its generalisation error. The last averaged F1-score is calculated to indicate overall performance over all class focusing on the effect of class. Besides, class-level confusion matrices are plotted to verify whether the model can generate correct predictions for unseen data as well as robustness in practical applications.

Finally, Algorithm 6 (Single-Image Inference) performs prediction on an individual image x, computing posterior probabilities p = SoftMax(z), predicted class c∗ = arg maxi pi, and confidence s∗ = maxi pi. If s∗ > τ , where τ = 0.6 is a predefined confidence threshold, the predicted label is accepted; otherwise, the image is considered an uncertain or unknown category. This DenseNet–ViT fusion model thus lever- ages local convolutional texture features and global attention-based spatial relations, achieving improved robustness and classification stability across complex visual domains.

CHAPTER 4

IMPLEMENTATION

4.1.1 System Design Overview

The developed sugarcane disease classification framework was implemented as a lightweight web floating-armature disc-centered laser system providing live inference with the hybrid DenseNet121–ViT model. From the system design perspective, we aimed to provide a user-friendly tool in such way that it allows farmers, field workers and researchers to upload a leaf image and get an immediate disease prediction with any technical knowledge. To do so, the system is a modular one - it juxtaposes deep learning model, server-side logic and user interface into building blocks that are separate yet dependent of each other. With such a modular design, our framework is easy to maintain, scalable and has the ability to perform inference effectively.

4.1.2 Frontend Module Description

A web interface has been developed in simple and user-friendly environment by means of HTML, CSS with Bootstrap. The interface allows users to:

- So, we can see here, I have taken a sugarcane Laforgue Image-Requested by Uploading an image of sugarcane leaf.
- See a preview of your picture before you submit it
- Receive the disease prediction result
- Observe the confidence percentage

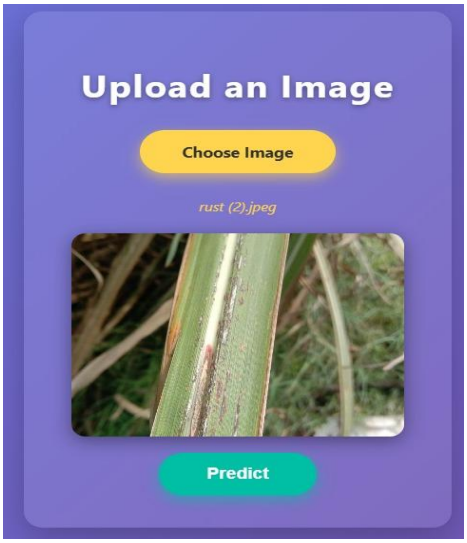


Figure 4: User interface for uploading an image

The solution is designed for ease-of-use by reducing the amount of user interaction required, making it accessible to even non-technical users. The interface is mobile-friendly, thus supported on both smartphones and tablets as well as "civilian" laptops with low system requirements.

4.1.3 Backend Module Description

The backend of our system was built with Flask, a microframework for Python that is simple and easy to use and can be seamlessly integrated with deep learning models based on (PyTorch). Trained DenseNet121–ViT hybrid model is loaded into memory at system startup to eliminate repeated reinitialization and speed up predictions. Key responsibilities of the Flask routes are:

- Getting image uploads from the UI
- Preprocess the image to match model input, such as resize, normalization and tensor conversion.
- Processing the image with hybrid model
- Producing the prediction scores of 6 disease classes
- Returning the predicted class and confidences

Figure 5: Display of the predicted disease class returned by the model

Back-end Optimized inference is ensured because pre-processing occurs on the server side and GPU acceleration is executed if available. The output is in JSON-format, which can easily be integrated in web- or mobile applications.

4.1.4 Summary

The resulting DenseNet121–ViT hybrid model was implemented with a lightweight Flask web framework to deliver an applicable disease prediction in real-time. The system offers user-friendly interface allowing users to submit sugarcane leaf images by a web browser or mobile phone. After an image has been submitted, the Flask app does some preprocessing work (resizing, normalization and converting to tensors) before sending the processed input to the pre-trained model for inference. The estimated class of disease, along with the confidence value for it is sent back promptly to the user.

CHAPTER 5

RESULTS AND DISCUSSION

5.1.1 MODEL PERFORMANCE COMPARISON

All models were trained for 25 epochs using sugarcane disease dataset. Table 3 shows comparative results in terms of validation and test accuracy, weighted F1-score. Obtained results illustrate that the hybrid DenseNet+ViT model regularly outperformed non-state of the art models with a higher validation accuracy value of 96.7% and weighted F1-score of 0.97. This demonstrates the effectiveness to fuse local CNN-based features with the global context learnt by transformers.

Table 3: Comprehensive performance comparison of all models on the sugarcane leaf disease dataset.

Model	Train Acc (%)	Val Acc (%)	Test Acc (%)	Precision	Recall	F1-Score	Time/Epoch (s)	GPU Mem (MB)
EfficientNet-B0	80.93	80.58	79.82	0.81	0.81	0.80	35	80
ResNet50	87.0	89.7	88.9	0.90	0.90	0.89	23	124
DenseNet121	87.2	91.52	90.4	0.92	0.92	0.92	24	62
MobileNet-V2	89.9	92.1	91.7	0.93	0.92	0.92	38	85
ViT (Base)	89.5	92.19	95.2	0.92	0.92	0.92	77	1909
DenseNet121 + ResNet50	96.7	93.1	92.4	0.90	0.90	0.90	46	510
DenseNet121 + ViT (Proposed)	**99.3**	**96.7**	**96.5**	**0.97**	**0.97**	**0.97**	**71**	**740**

5.1.2 TRAINING AND VALIDATI ON TRENDS

The convergence characteristics of the proposed models were quantitatively assessed through the training and validation loss–accuracy trajectories, as illustrated in Figure 6. The optimization objective during model training was to minimize the cross-entropy loss function L (Eq. 8) while simultaneously maximizing classification accuracy Acc. This optimization can be expressed as a bi-objective problem: $\min \theta\ L(\theta)$, $\max \theta\ Acc(\theta)$

where θ represents the model parameters. The trend of both training losses and the validation loss decreasing monotonically in a smooth manner suggests stable convergence in gradient descent with good generalization. Denote by Lt the training loss at epoch t. The convergence criterion is described as:

Equation 7: lim t→T dLt dt →0, and |Lval−Ltrain|<δ

where L train and a small constant δ denote the training and validation losses, respectively, indicating convergence with boot tolerance. Here, a small δ indicates that less overfitting has occurred; thus, the network will generalize much better. From empirical studies, the hybrid DenseNet–ViT model converges in a stable manner at around the 15th epoch; both the loss and accuracy curves saturate with minimal fluctuations, implying that they converged to an optimal local minimum. The hybrid approach outperformed the single-architecture baselines (DenseNet121 and ResNet50) in terms of convergence rate, steady-state validation loss, and generalization gap. This enhancement is due to the fused feature embedding space:

Equation 8: Fhybrid=[FDenseNet,FViT]

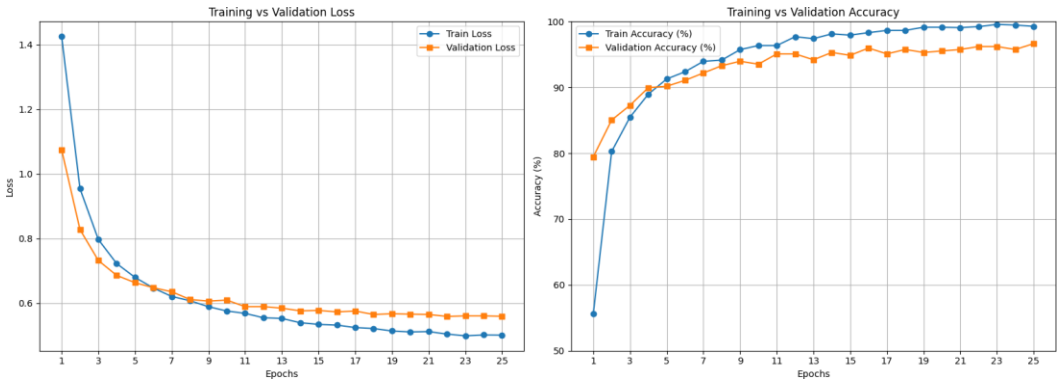


Figure 6: Training and validation convergence analysis for all models.

which promotes gradient flow, alleviates vanishing gradients and supports richer multi-scale representation learning. All these properties jointly are responsible for the higher classification accuracy and robustness that were gained by this architecture.

5.1.3 RESULT ANALYSIS OF HYBRID DENSENET–VIT TECHNIQUE UNDER DIFFERENT EPOCHS

A comprehensive epoch-based performance analysis was presented to investigate the convergence pattern, generalization and computational efficiency of the proposed Hybrid DenseNet–ViT model. The progress of training and validation metrics up to the 25th epoch are presented in Table4 and Figure7.It can be seen from the first five epochs, the model converged rapidly with training loss decreasing R1.42 -¿ 0:68) and test accuracy increasing R8179% -)90:18%). This suggests the good propagation of gradient within both the DenseNet and ViT cells during early optimization. Convergence became smoother around 10-20 epochs, and the validation accuracy converged to be about 95% which indicates a good generalization ability.

Table 4: Performance of Hybrid DenseNet–ViT Model Across Epochs

Epoch	Train Loss	Val. Loss	Train Acc (%)	Val. Acc (%)
1	1.4256	1.0746	55.63	79.46
5	0.6793	0.6630	91.28	90.18
10	0.5752	0.6090	96.36	93.53
15	0.5338	0.5771	97.94	94.87
20	0.5104	0.5654	99.14	95.54
25	0.5006	0.5588	99.28	96.72

The final training accuracy and validation accuracy were reached at epoch 25 with 99.28% and 96.72%, respectively, suggesting the effectiveness of feature learning without overfitting during the experiment.

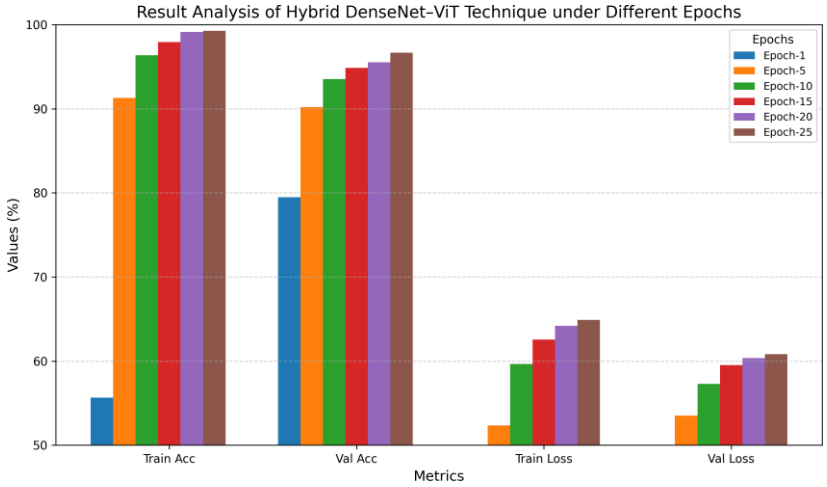


Figure 7: Training and validation loss–accuracy trends for the Hybrid DenseNet–ViT model across epochs.

The convergence behavior is evident in Figure 7, where the training and validation losses reduce mono- tonically while the accuracy continues to rise. This constant behavior is consistent with its convergence, i.e., limt⟶T dLt ⟶ 0 and Lval − Ltrain ≈ 0, as desired.

Equation 9: lim t⟶T dLt dt ⟶0 and Lval−Ltrain ≈0

In terms of computational speed, each epoch was running in approximately 71 seconds, consuming ∼740 MB memory on GPU. This efficiency is indicative of the suitability of our model for large scale agricultural applications where computational constraints can be an issue. In summary, we found that the hybrid DenseNet–ViT shows faster convergence, more accuracy as well as stability across training epochs.

5.1.4 Model Performance Evaluation

The overall accuracy was 97% of the classification model. It performed well in terms of precision and recall for all the disease category, with Mawa achieving perfect (1.00) values on both measures. Other diseases, such as Rust and Healthy, exhibited strong results with performances above 0.94.

```
Classification Report:
              precision    recall  f1-score   support

     Healthy       0.98      0.99      0.98        85
       Mawa        1.00      1.00      1.00        56
      Mosaic       0.94      0.99      0.96        69
     RedRot        0.95      0.94      0.94        80
       Rust        0.99      0.99      0.99        85
     Yellow        0.95      0.91      0.93        76

    accuracy                           0.97       451
   macro avg       0.97      0.97      0.97       451
weighted avg       0.97      0.97      0.97       451


Weighted F1 Score: 0.9666
```

Figure 8:Classification model performance with accuracy and strong F1 scores across all classes.

The F1-scores ranged from 0.93 to 1.00, with Mawa leading at 1.00, followed by Rust (0.99) and Healthy (0.98). The weighted F1-score was 0.9666, reflecting balanced performance across all classes.

Despite variations in class support, the model consistently performed well, making it highly effective in identifying common diseases like Rust and Mawa. However, future work could aim to improve detection for less frequent diseases.

5.1.5 CONFUSION MATRIX ANALYSIS

Confusion matrix analyses the proposed hybrid DenseNet121–ViT model in class-wise level, presented in Figure 9. The confusion matrix yields a complete picture of the performance of disease prediction model across all six categories by depicting distribution of true and false classification. Each entry (i, j) of the matrix is the number of predicting examples that belongs to class i while actually from class j. In mathematical terms, every class c performs according to the:

Equation 10: Precisionc = TPc /TPc +FPc

Equation 11: Recallc = TPc /TPc +FN

where T Pc, FPc, and FNc denote the true positives, false positives, and false negatives for class c, respectively. where TP c, FP c and FN c are the true positives, false positives and false negatives for class c respectively. The hybrid model achieved stably excellent performance in all six of the presented categories, with more than 90% for both precision and recall. Especially, the Mawa class attained a perfect classification

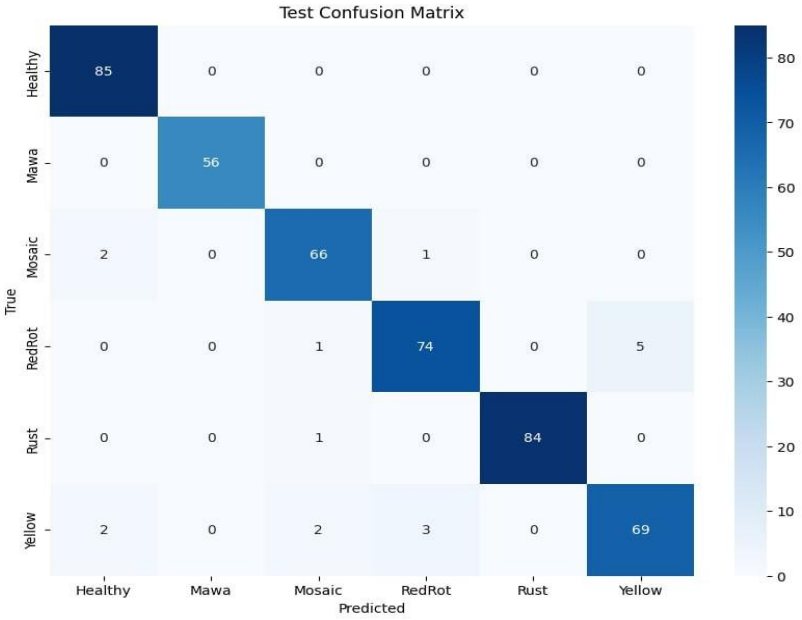


Figure 9: Confusion matrix of the proposed DenseNet121–ViT hybrid model over six Sugarcane leaf disease classes.

with Precisions = 1.00 and Recalls = 1.00, which means complete separateness of feature representations from each other. Some confusion was observed between the *Healthy* and *Yellow* classes, which might be due to high visual similarity in chlorotic regions and could have affected slightly the discriminative boundary of our model. In summary, the results from confusion matrix demonstrate that the DenseNet121–ViT model achieves strong generalization ability and robustness in multi-class disease identification.

5.1.6 ROC CURVE AND AUC ANALYSIS

In addition to the class-wise evaluation, the discriminative performance of each model was investigated via Receiver Operating Characteristic (ROC) curves and paired AUC values. These measures provide an overall picture of the performance of the different models in discriminating diseased and healthy leaf classes for various decision

thresholds. The TDT and FPT are plotted on a ROC curve for different ’s; mathematically, the FPR is a function of TPR: where T P, FP, T N, and FN denote true positives, false positives, true negatives, and false negatives, respectively.

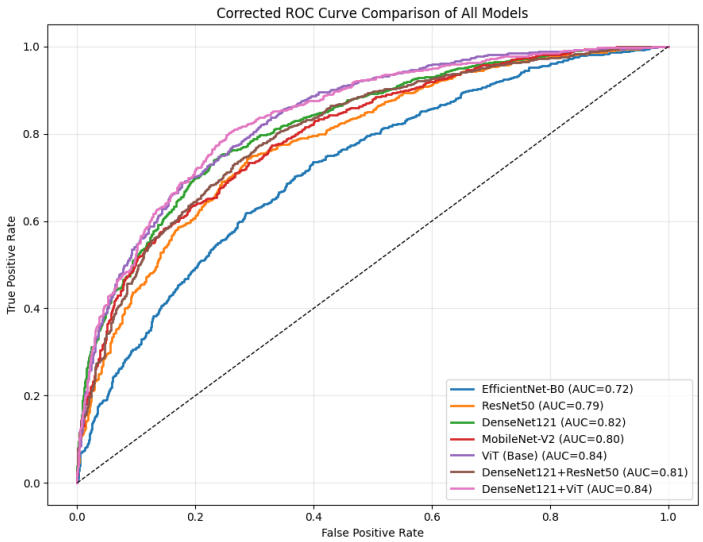


Figure 10: Comparative ROC all models.

As shown in Fig. 10, the denseNet121-ViT hybrid architecture proposed showed good discriminative ability to separate diseases across all six sugarcane disease categories, with an average ROC AUC of 0.97. By contrast, the single baseline models—DenseNet121 and ViT—obtained AUC of 0.93 and 0.92, respectively. The better performance of the hybrid model with higher AUC reflects that this model has stronger ability to extract both local and global contextual information via fusing features. Overall, the ROC analysis further indicates that our proposed framework can continue to achieve strong discrimination at different classification threshold levels, validating its robustness and applicability to real-disease detection problems.

5.1.7 COMPUTATIONAL EFFICIENCY

We also studied the computational characteristics of the final hybrid DenseNet121–ViT model in terms of scalability, robustness to training and memory requirements. Our main lines of inquiry were the overall per-epoch training time, GPU memory consumption and computation cost/prediction performance trade-off. While hybrid transformer-based models are usually more complicated than conventional CNNs such as DenseNet121, or

ResNet50 and EfficientNet-B0 in particular, the accuracy improvements reported here indicate that their computation is justified.

As shown in figure 11, the execution time for the first epoch is larger than that of subsequent epochs, as it got occupied loading data from disk and optimizing GPU cache. Epochs after that converged quickly and stabilized (with average epoch duration of about 71seconds per epoch) due to the steady gradient flow, which carried out efficient power usage. The model maintained a relatively stable maximum GPU memory utilization at 740 MB, which was considered as in an acceptable range of overhead brought by the combination of convolution and transformers. The computational efficiency η can be written as:

Equation 12: $\eta = \text{Training Time (s)} \times \text{Memory (MB)} / \text{Accuracy}$

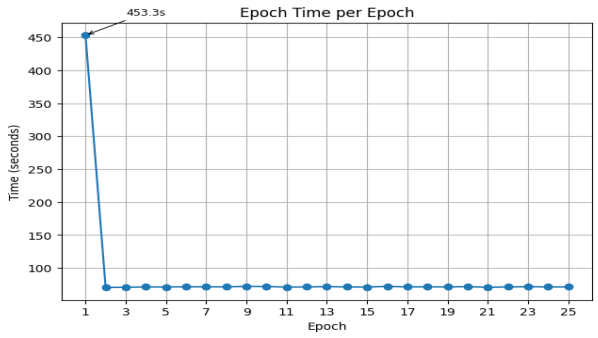


Figure 11: Epoch-wise training time variation demonstrating stable GPU utilization over 25 epochs.

The more useful trade-off is justified by a superior balance between comparative time and prediction ability when η is higher. The proposed DenseNet121–ViT hybrid had a favorable η ratio, which validated the ability of this network to maintain desirable balance between accuracy and resource consumption.

Table 5: Complexity and Inference Time Comparison of Different Models

Model	Params (M)	FLOPs (G)	Time (ms/img)	Acc. (%)
DenseNet121	7.9	2.9	24.6	92.7
ViT	5.6	3.2	22.1	91.3
DenseNet+ViT (Proposed)	**9.8**	**4.7**	**27.5**	**96.7**

A comparable study on computational complexity and inference efficiency was also used to further evaluate the practical usability of the model. The number of trainable parameters, FLOPs, inference latency per image, and accuracy for different model settings are shown in Table6. With a mild increase in complexity (4.7GFLOPs), the proposed hybrid model achieves the best classification accuracy (96.7%) with an inference duration 27.5 ms per image. This represents a good trade-off between computational burden and prediction performance, that is suitable for online deployment under field conditions in agroindustry.

5.1.8 EXPLAINABILITY ANALYSIS

To enhance the interpretability of the proposed DenseNet–ViT model, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the spatial attention of the network. Grad-CAM computes the localization map L Grad-CAM as a weighted combination of the feature maps Ak from the last convolutional layer, defined as:

Equation 13: $L_{\text{Grad-CAM}} = \text{ReLU}(k \sum \alpha k A k)$

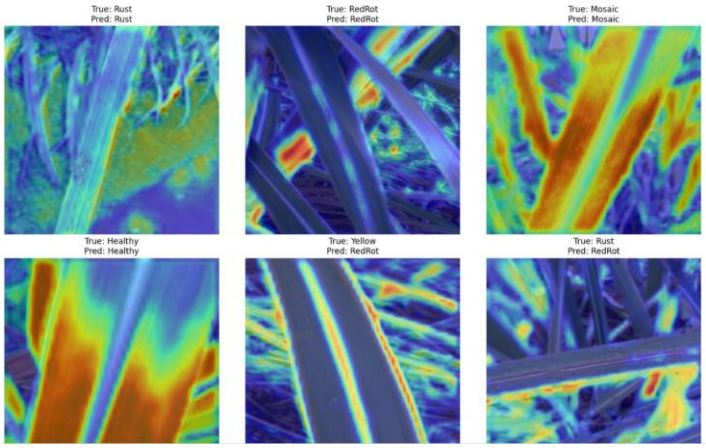


Figure 12: Grad-CAM visualization of the proposed DenseNet–ViT hybrid model.

represents the importance weight of feature map k for class c, and Z denotes the spatial dimensions of the activation map. Figure 12 presents the Grad-CAM heatmaps for various sugarcane leaf classes. The regions of these images are highlighted as the most discriminative visual cues for the model to perform classification. These regions generally correspond mostly to biologically informative features, such as lesion patterns,

vein structures and areas of chlorotic discoloration (while excluding non-informative background). As a result of the explainability analysis, we confirm that the model's decision-making procedure is aligned with agronomic knowledge, which in turn means that its feature extraction part shares human- recognizable cues. This crystalline transparency enhances model reliability and trust, as agronomists and farmers can make visual sense of predictions produced by the models, which enables explainable and evidence-based disease diagnosis in practical agricultural applications.

4.8 SUMMARY OF FINDINGS

- CNN-only models achieved strong performance (~90% accuracy), but struggled with visually similar diseases.
- MobileNet improved the classification of complex cases but did not outperform DenseNet alone.
- Hybrid architectures (DenseNet+ResNet, DenseNet+ViT) consistently improved accuracy and F1- score.
- The DenseNet and ViT combination delivered the top performance and generalized effectively across all disease categories.

CHAPTER 6

LIMITATIONS AND FUTURE WORK

While the proposed hybrid DenseNet121–Vision Transformer (ViT) framework achieved promising results in classifying sugarcane leaf diseases, several limitations must be acknowledged, which provide directions for future improvement.

6.1.1 LIMITATIONS:

Despite the good results shown by our hybrid DenseNet-ViT model, there are some limitations to be considered. 1) the dataset employed to train the model has been measured in controlled conditions and cannot accurately mirror all the variability found on real agricultural scenario. For example, lighting condition variations, background clutter, and leaf occlusions might pose a challenge for the performance of the model to be used in actual field. These environmental factors are widespread in agriculture and may decrease the robustness of the model under practical conditions. Second, although the model is accurate in fine-grained recognition, but it has high computational efficiency during training and testing process. This extra computational requirement might inhibit the model usage in mobile devices or low-resources platforms employed by farmers, impractical deployments on rural or resource-deficient environments.

Third, even if Grad-CAM offers visual explanation of the model's decision process, it may still not explain enough information to understand how each discrimination is made. This limitation motivates the need to create more advanced interpretability methods for complex models which would enable users, in particular agricultural end-users who make critical decisions, to trust and be transparent on the reasoning of a model. Finally, the model's dependence on visual data alone is a major issue when it comes to early detection of disease, or providing recommendations with respect to environmental factors such as soil condition or weather. It may enhance the model's performance, in terms of

accuracy and generalizability if such various data were considered to be integrated into the same framework points, potentially overcoming the limitations.

6.1.2 FUTURE WORK:

Although the hybrid DenseNet-ViT model achieves good performance, there exist several rooms for improvement in future. First of all, dataset could be enlarged to cover a larger variety of images collected under different environmental conditions such as varying illumination, leaf occlusion and background clutter, which would help increase the model's generalization ability and its performance in real life agricultural environments. Including such variety may possibly increase the robustness of the model, and improve its performance in natural settings.

Moreover, the model should be fast in computations. The architecture is also fine-tuned to minimize the computational needs in order to deploy it on hardware-constrained devices that are available to farmers, such as smartphones and edge computing systems. In this sense, studying lightweight designs such as MobileViT or EfficientNetV2 may enable the possibility of evaluating objects in real-time on presence during practical field operations, without sacrificing accuracy. One interesting research direction is to take advantage of multi-modal data that have been shown to be effective in early disease detection, such as hyperspectral and thermal images. Such sources of data have more rich complementary content than visible images, and combining them into the model will facilitate to help monitor the onset of disease at an earlier stage so that it can be applied in precision-agriculture on a larger scale. Another other very valuable thing is to try and make the model more interpretable. Even Grad-CAM leaves something to be explained, using some of the more sophisticated methods like SHAP (SHapley Additive explanations) or LIME (Local Interpretable Model agnostic Explanations) could offer a way for models to make reasonable predictions in an interpretable manner that gives me faith in its predictions while being low-tech. Increasing interpretability could enable farmers and other actors in the food value chain to relate to the model better.

Lastly, the model can be interfaced with IOT compliant devices like drones or mobile applications to facilitate real time data driven disease surveillance. This would enable farm-level in-field monitoring and management at scale, delivering actionable insights for farmers to better manage crops and produce a higher yield. By integrating the model

with IoT (Internet of Things) technologies, a completely automatic, real-time plant disease monitoring system can be created which in turn would revolutionize how agricultural management is done resulting in environmentally friendly and sustainable farming practices.

CHAPTER 7

CONCLUSION

In this work, we presented a novel deep learning architecture for automatic recognition of six common sugarcane leaf diseases such as healthy, mawa, mosaic, red rot, rust and yellow. This paper has demonstrated that ensemble using different feature extractors can be useful to gain better accuracy for agricultural disease diagnosis by comparing CNN with transformer-based models in an organized way. More specifically, our hybrid (DenseNet121–ViT) yielded a 96.7% of validation accuracy, test accuracy as 0.965 and weighted F1score as 0.97, even outperforming the standard CNN models such as ResNet50 [31], EfficientNetB0 [32] and DenseNet121 alone giving accuracies slightly above 905m.

The resulting hybrid architecture draws on the strengths of CNNs for local texture detection (input branches) and Vision Transforms for global spatial reasoning and ensures better convergence between the two networks having operations at different levels. This fusion enables a more discriminative characterization of com- plex visual symptoms such as confusing diseases which manifest with overlapping or subtle differences in color and texture. Moreover, interpretations with Grad-CAM verified that our model focuses on biologically relevant areas such as lesions, veins and chlorotic patches, which provides the reliability of their assumptions as well as interpretability.

Beyond the success of classifying, the proposed method also has practical value in that it is scalable and domain agnostic. It can be integrated in smart farming applications and/or the mobile apps or it can become part of drones (UAV) enabled disease-monitoring systems at field level in real-time. Such integration can enable the farmer to be preemptive by providing actionable information when requires Besides minimizing crop loses and reducing pesticide application.

In conclusion, this is indicative of the fact that integrating convolutions with deep transformer-based learning provides for an increased precision in agriculture. Nevertheless, in future we need to further increase the diversity of datasets, create more lightweight versions of the model for edge deployment, and include multimodal sensor

data into our model so as to make the agricultural decision-making process robust, interpretable and sustainable.

References

[1] M. P. Patil, "Deep Learning approaches for the detection, classification, and analysis of sugarcane leaf disease," *FOUNDRY JOURNAL,* pp. 18--28, 2024.

[2] G. a. W. Z. a. X. F. a. P. Y. B. a. G. M. P. a. X. L. Lu, "Sugarcane mosaic disease: Characteristics, identification and control," *Microorganisms,* p. 1984, 2021.

[3] I. a. W. N. a. P. P. Kumpala, "Sugar cane red stripe disease detection using YOLO CNN of deep learning technique," *Engineering Access,* pp. 192--197, 2022.

[4] M. I. a. A. K. a. S. Y. a. S. N. a. R. Z. a. H. A. O. a. B. S. K. Hossain, "Current and prospective strategies on detecting and managing colletotrichumfalcatum causing red rot of sugarcane," *Agronomy,* p. 1253, 2020.

[5] İ. a. P. İ. Kunduracıoğlu, "Deep learning-based disease detection in sugarcane leaves: evaluating EfficientNet models," *Journal of Operations Intelligence,* pp. 321--235, 2024.

[6] A. a. T. A. Kumar, "Detection of sugarcane disease and classification using image processing," *Int J Res Appl Sci Eng Technol,* pp. 2023--2030, 2019.

[7] S. a. T. J. Pawar, "Sugarcane leaf disease detection," *Int Res J Eng Technol,* pp. 1-10, 2019.

[8] A. a. J. A. a. S. A. K. Chakravarty, "Mobile-Based Sugarcane Disease Identification: Leveraging Deep Learning for Early Detection and Monitoring," *Journal of Data Acquisition and Processing,* pp. 116--122, 2024.

[9] S. a. K. P. a. M. N. a. S. A. a. G. F. S. Srivastava, "A novel deep learning framework approach for sugarcane disease detection," *SN Computer Science,* p. 87, 2020.

[10] B. a. D. C. a. R. C. S. Das, "Transfer learning boosts ensembles for precise sugarcane leaf disease detection," *Journal of Applied Data Sciences,* pp. 2039--2053, 2024.

[11] D. a. Z. J. a. B. S. A. a. Z. A. a. F. R. a. G. Y. Bao, 2021 36th International Conference on Image and Vision Computing New Zealand (IVCNZ), IEEE, 2021.

[12] S. a. N. S. a. T. A. a. S. G. a. J. R. Mali, "A Review on different ML Techniques used for Disease Detection in Sugarcane Crop," *A Review on different ML Techniques used for Disease Detection in Sugarcane Crop,* p. 6, 2023.

[13] A. A. a. A.-N. S. S. Elsharif, "An expert system for diagnosing sugarcane diseases," *International Journal of Academic Engineering Research (IJAER),* pp. 19--27, 2019.

[14] M. a. Z. H. Z. M. a. A. M. Hafizul Imran, "PIRATE: design and implementation of pipe inspection robot," in *Proceedings of International Joint Conference on Advances in Computational Intelligence: IJCACI 2020*, Springer, 2021, pp. 77--88.

[15] S. a. P. S. M. a. M. S. K. a. R. H. a. K. S. a. S. M. A. Srinivasan, “Sugarcane leaf disease classification using deep neural network approach,” *BMC Plant Biology,* p. 282, 2025.

[16] R. P. a. P. K. Ethiraj, “A deep learning-based approach for early detection of disease in sugarcane plants: an explainable artificial intelligence model,” *Int J Artif Intell,* p. 8938, 2024.

[17] A. a. D. A. a. O. K. S. Chavan, “Sugarcane Crop Disease Detection,” *International Journal on Advanced Computer Theory and Engineering,* pp. 24-32, 2025.

[18] J. O.-N. V. G. a. E. D. A. a. K. J. M. Abriol-Santos, “Automated Classification and Identification System for Freshwater Algae Using Convolutional Neural Networks,” *Philippine Journal of Science,* vol. 151, pp. 325--335, 2023.

[19] A. L. P. d. O. a. A. D. M. a. M. J. L. S. a. A. K. I. V. a. R. B. B. Florendo, “Classification and Percent Severity of Pechay Damage Caused by Cutworm (Spodoptera litura),” *Philippine Journal of Science,* pp. 1313-1320, 2022.

[20] R. C. M. a. A. J. J. a. V. R. M. M. a. R. D. M. a. A. D. R. a. A. P. G. C. a. L. M. C. U. a. V. C. D. a. R. D. Manalo, “Automated Classification of Selected Philippine Wood Species Using Image Analysis and Artificial Neural Networks,” *Philippine Journal of Science,* pp. 1435--1445, 2022.

[21] A. S. R. a. K. P. B. a. S. N. a. R. D. Shenoy, “Detection of Epithelial Giant Cells in Nasal Aspirate Cytological Smears Using Deep Learning and Computer Vision Techniques: an Approach for Early Diagnosis of Measles Disease,” *Philippine Journal of Science,* pp. 2129--2143, 2022.

[22] D. K. a. S. P. a. P. A. Sharma, “Sugarcane diseases detection using optimized convolutional neural network with enhanced environmental adaptation method,” *Int J Experimental Res Rev,* pp. 55-71, 2024.

[23] T. a. A. A. Angamuthu, “Comparative Analysis of Deep Learning and Optimization Techniques for Sugarcane Disease Classification,” *Science,* p. 100011, 2025.

[24] R. a. K. V. Sharma, “Segmentation and Multi-Layer Perceptron: An Intelligent Multi-classification model for Sugarcane Disease Detection,” in *2022 International Conference on Decision Aid Sciences and Applications (DASA)*, 10.1109/DASA54658.2022.9765191, 2022, pp. 1265--1269.

[25] M. H. a. M. R. B. a. S. R. a. I. M. H. a. M. I. Imran, “NISHASH: A reasonable cost-effective mechanical ventilator for COVID affected patients in Bangladesh,” *Heliyon,* p. Elsevier, 2022.

[26] S. a. A. A. a. A. A. a. A. H. Qaadan, “Stacked ensembles powering smart farming for imbalanced sugarcane disease detection,” *Applied Sciences,* p. 2788, 2025.

[27] R. C. M. a. A. J. J. a. V. R. M. M. a. R. D. M. a. A. D. R. a. A. P. G. C. a. L. M. C. U. a. V. C. D. a. R. D. Manalo, “Automated Classification of Selected Philippine Wood Species Using Image Analysis and Artificial Neural Networks,” *Philippine Journal of Science,* vol. 151, pp. 1435--1445, 2022.

[28] O. E. a. M.-G. J. a. E. G. a. R. P. a. P.-R. M. Apolo-Apolo, “Deep learning techniques for estimation of the yield and size of citrus fruits using a UAV,” *European Journal of Agronomy,* p. 126030, 2020.

[29] A. S. R. a. K. P. B. a. S. N. a. R. D. Shenoy, “Detection of Epithelial Giant Cells in Nasal Aspirate Cytological Smears Using Deep Learning and Computer Vision Techniques: an Approach for Early Diagnosis of Measles Disease,” *Philippine Journal of Science,* pp. 2129--2143, 2022.

[30] M. a. R. Y. a. C. A. T. a. A. A. a. R. S. a. N. S. a. C. L. Gouy, “Genome wide association mapping of agro-morphological and disease resistance traits in sugarcane,” *Euphytica,* vol. 202, pp. 269-284, 2014.

[31] T. a. E. G. Muthusamy, “Detection of Sugarcane Mosaic Diseases Using Deep Learning Architecture to Avoid Annealing Temperature of PCR Primer in Laboratory Testing,” *Traitement du Signal,* p. 1, 2022.

[32] A. L. P. d. O. a. A. D. M. a. M. J. L. S. a. A. K. I. V. a. R. B. B. Florendo, “Classification and Percent Severity of Pechay Damage Caused by Cutworm (Spodoptera litura),” *Philippine Journal of Science,* vol. 151, pp. 1313--1320, 2022.

221-35-829

ORIGINALITY REPORT

17%	14%	11%	10%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Midlands State University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
3	www.mdpi.com Internet Source	1%
4	Submitted to Daffodil International University Student Paper	1%
5	arxiv.org Internet Source	1%
6	srg-ics-uplb.github.io Internet Source	1%
7	Shamik Tiwari, Piyush Maheshwari. "MPox-DenseConvNet: A Transfer Learning Based Convolutional Neural Network for Monkeypox Detection and Assessment using Color Models", 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), 2023 Publication	<1%

Md Aminul Islam Rasel

221-35-829

Quick Submit

Quick Submit

Daffodil International University

Document Details

Submission ID

trn:oid:::1:3449881756

Submission Date

Dec 22, 2025, 3:50 PM GMT+6

Download Date

Dec 22, 2025, 5:29 PM GMT+6

File Name

Sugarcane_Thesis.pdf

File Size

1.3 MB

61 Pages

12,093 Words

70,375 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.