DETECTION OF BANANA FUSARIUM WILT & BLACK SIGATOKA: A DEEP LEARNING APPROACH FOR SMALLHOLDER FARMS IN CENTRAL UGANDA

PETER MULINDWA

M23M19/252

A DISSERTATION SUBMITTED TO THE FACULTY OF ENGINEERING, DESIGN AND TECHNOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF SCIENCE IN DATA SCIENCE AND ANALYTICS OF UGANDA CHRISTIAN UNIVERSITY

June,2025

UGANDA CHRISTIAN UNIVERSITY
A Centre of Excellence in the Heart of Africa

ABSTRACT

Despite bananas being one of the most important sources of food and revenue for farmers in central Uganda, banana cultivation is severely threatened by destructive diseases such as yellow and black Sigatoka and Fusarium wilt, also known as Panama disease. Smallholder farmers, who are the backbone of Uganda's agricultural sector, often rely on manual inspection methods to identify these diseases, which are time-consuming, prone to error, and highly ineffective for early disease interventions.

For an accurate, efficient, and scalable banana leaf disease detection method (specifically Fusarium wilt and black Sigatoka), this thesis proposes a hybrid deep learning image classification approach which integrates Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Gray Level Co-occurrence Matrix (GLCM) extracted texture features.

To carry out this research, a dataset of approximately 17,000 annotated banana leaf images divided into three classes (healthy leaves, Fusarium wilt and black Sigatoka) was obtained from the Lacuna Banana Project. To enhance the robustness of the model, the image dataset was rigorously preprocessed by resizing, normalising, and augmenting the images. With the help of GLCM, texture features were extracted, and these were combined with spatial features learnt by the CNN and ViT models to improve classification sensitivity. A number of deep learning models were developed and evaluated; these included a custom CNN, transfer learning with InceptionV3, a ViT-based architecture and a Hybrid model. To assess the model 's performance, metrics such as accuracy, precision, recall, and F1-score were used.

After evaluating all the trained models, the Vision Transformer (ViT) model outperformed all the other models with a classification accuracy of 99%, while the proposed hybrid model achieved a balanced accuracy of 98%, with significant precision and recall across all disease categories. The integration of GLCM features greatly improved the detection of texture-specific diseases such as black sigatoka. This research contributes to a robust, interpretable, and field-deployable AI-based diagnostic tool that aligns with Uganda's national goals for data-driven agricultural development and the Sustainable Development Goals related to food security and sustainable agriculture.

DECLARATION

I, Peter Mulindwa, hereby declare that this thesis report (Detection of Banana Fusarium Wilt & Black Sigatoka: A Deep Learning Approach for Smallholder Farms in Central Uganda) is my original work and has not been submitted to any other institution of higher education for the award of a degree or diploma.

Following academic integrity standards, acknowledgements and citations have been made where other scholars' works have been used. This research reflects my efforts and understanding and was carried out under the guidance and supervision of my academic supervisors at Uganda Christian University. I take full responsibility for the content and any errors that may be found in this work.

Name: Peter Mulindwa
Registration Number: M23M19/252
Date: 9^{th} June 2025

Contents

List of Figures

List of Tables

List of Acronyms

Acronym	Full Meaning
AI	Artificial Intelligence
CNN	Convolutional Neural Network
ViT	Vision Transformer
GLCM	Gray Level Co-occurrence Matrix
SDG	Sustainable Development Goal
NDP	National Development Plan
NPA	National Planning Authority
TARI	Tanzania Agricultural Research Institute
ReLU	Rectified Linear Unit
ROI	Region of Interest
HSV	Hue Saturation Value
GPU	Graphics Processing Unit
API	Application Programming Interface
p	Probability
ROI	Region of Interest
PIAP	Programme Implementation Action Plan
BXW	Banana Xanthomonas Wilt

1.0 INTRODUCTION

1.1 BACKGROUND OF STUDY

According to Sheridan, a botanist at UCLA, bananas (Musa acuminata) are a fruit formed by multiple ovaries in a single flower, characterised by their elongated shape, yellow skin, and creamy sweet flesh. As a staple food for millions worldwide (Sheridan, 2019, p. 12), bananas hold a special place in Ugandan livelihood, particularly in the Central and Western regions, where they are a major cash crop. Farmers attest that banana cultivation has improved their livelihood and fostered food security, with 75% adapting to its growth.

Bananas also have a rich history and cultural importance in Uganda. For example, in Buganda culture, banana species such as Bogoya, Ndizi, Ganja, and Kivuvu are essential to both traditional diets and formal occasions, according to Nakayiwa (2020).

Despite all these benefits, banana production is threatened by diseases such as Xanthomonas wilt (BXW), commonly known as Banana Bacterial Wilt, Fusarium wilt, also known as Panama disease, and Sigatoka. Nakayiwa and Kigenyi (2019) further emphasise the magnitude of the impact of these diseases on banana productivity since they cause vascular damage, reduced yields, soil contamination, root rot, food insecurity, and economic depreciation. Early detection is necessary, and this project uses deep learning models such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to facilitate banana disease detection by providing accurate, scalable, and timely solutions to farmers and agricultural stakeholders in central Uganda.

According to Krizhevsky (2012), CNNs achieve object detection and image classification through the use of activation functions, pooling layers, dense layers, convolutional layers, and flattened layers. This method is perfect because CNNs can handle large images, are resilient to changes, and can extract features.

1.2 Problem Statement:

There is a serious threat of banana disease prevalence, such as Fusarium wilt, Xanthomonas wilt, and Black Sigatoka, among smallholder farms in central Uganda. These diseases lower yields and incomes, in addition to causing food insecurity. Early detection is essential to stop these diseases from spreading and lessen the harm they cause. The current detection methods are labour-intensive, time-consuming, and prone to error because they heavily rely on visual inspections by farmers and extension agents.

A lot of research has been undertaken to address this issue (Nakayiwa et al., 2018, 2019); however, the diseases have intensified, severely affecting the farmers' livelihoods and productivity. To safeguard banana production and protect the livelihoods of smallholder farmers, the challenges resulting from traditional methods, which call for hand field inspections, highlight the need for a more efficient, exact, and scalable means of identifying and treating these diseases.

1.3 Justification of the Study

This research justifies the pressing need to maximise artificial intelligence, especially deep learning algorithms, to solve the significant threat of banana diseases endangering smallholder farms in Central Uganda. This work closes current gaps in banana disease management by developing a hybrid deep learning model combining CNNs, Vision Transformers, and Grey Level Co-occurrence Matrix (GLCM) features to detect banana diseases from leaf images, thus improving productivity and reducing the terrible impact of these diseases.

The justification for this study is multifaceted; that is to say, this research leverages AI advancements to improve agricultural productivity and food security; it also addresses the limitations of traditional disease detection methods, which are time-consuming, labour-intensive and prone to inaccuracies; last but not least, this study is helping the farming sector to be sustainable, in line with the Sustainable Development Goals (SDGs), by offering a fast diagnosis tool for farmers and agricultural stakeholders, enabling timely interventions and improved disease management.

By accurately identifying banana diseases from leaf images, this research has the potential to transform the livelihoods of smallholder farmers and ensure a more secure and sustainable food supply for communities in Central Uganda.

1.4 Research Questions

1. How effective are CNNs and Vision Transformers in detecting banana diseases from leaf images?
2. In what ways do GLCM features improve the accuracy of CNN and ViT models for disease detection?
3. What is the optimal combination of CNN and ViT architectures to achieve the highest accuracy for banana disease detection?

1.5 Hypothesis:

1. Deep learning models (CNNs and ViTs) will outperform traditional visual inspection methods in detecting banana diseases.
2. The addition of GLCM features will enhance the model's ability to detect texture-based diseases such as Black Sigatoka.
3. Fine-tuned pre-trained models (InceptionV3 & ViTs) will perform better than CNN models trained from scratch, providing higher accuracy and efficiency.

1.6 Main Objective:

To propose and validate a scalable deep learning approach for the early detection and classification of Fusarium wilt and Black Sigatoka in banana plants in Central Uganda, thereby improving crop productivity and food security.

1.6.1 Specific objectives:

i. To design and develop a deep learning model architecture that integrates CNNs, Vision Transformers, and GLCM features to enhance classification accuracy.

ii. To evaluate the performance of the developed models using metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness.

iii. To compare the performance of the hybrid model against a custom CNN, and a ViT-based architecture to demonstrate its superiority.

iv. To optimise the model by fine-tuning hyperparameters and exploring different architectures to improve performance.

2.0 Literature Review

This literature review explores the body of research on disease detection, specifically bananas. Emphasis is placed on Deep learning approaches such as Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), Hybrid models, and texture analysis methods such as Gray-Level Co-occurrence Matrix (GLCM). In this chapter, I will evaluate the techniques and methodologies previously applied, highlighting their strengths and weaknesses in addition to establishing their applicability in central Uganda.

2.1 Overview of Common Banana Diseases

Banana diseases such as Fusarium Wilt (Tropical Race 4), Black Sigatoka, and Xanthomonas Wilt lead to significant and persistent challenges in the agricultural sector, leading to substantial crop yield losses and therefore threatening food security and the economic stability of farmers who rely on bananas for sustainability.

In Uganda, these pathogens reduce yields by approximately 30 to 80% through vascular destruction and leaf necrosis as documented by (Rietveld et al., 2014)

This review provides an overview of the main banana diseases that affect smallholder farms in central Uganda, and these include;

Xanthomonas Wilt (Banana Bacterial Wilt). Caused by the bacterium Xanthomonas campestris pv. Musacearum is characterised by the wilting and yellowing of leaves (Tripathi et al., 2009).

Black Sigatoka. Caused by Mycosphaerella fijiensis and it's a fungal disease distinguished by leaf streaks and black spots, which reduce photosynthetic ability (Marin et al., 2003).

Yellow Sigatoka. caused by Mycosphaerella musicola, and it is characterised by yellow leaf streaks which affect plant vigour and yield (Marin et al., 2003).

2.2 The Evolution of Disease Detection

Over the years, plant disease detection has evolved from traditional manual inspection to advanced AI-driven methodologies. Early automation methods employed traditional image processing techniques such as spectral feature analysis, threshold segmentation, Canny edge detection algorithms, and histogram analysis(Upadhyay et al., 2025). Although these methods proved to be helpful, they often faced challenges in effectively handling complex image backgrounds and adapting to multiscale disease features, which are common in agricultural settings(Xiao et al., 2025).

As years passed, classical machine learning algorithms such as Support Vector Machines (SVMs), Naïve Bayes, Decision Trees, k-Nearest Neighbors (KNN), Random Forests, and Multilayer Perceptrons (MLPs) gained popularity(Mehdipour et al., 2025). These algorithms facilitated the automation of disease diagnosis, but frequently struggled with scalability since their effectiveness was heavily based on handcrafted feature engineering, which required domain expertise(Mehdipour et al., 2025).

The arrival of deep learning algorithms marked a transformative phase in computer vision and its application in precision agriculture. Deep learning models such as Convolutional Neural Networks can automatically learn hierarchical representations of image data, effectively bypassing the need for manual feature engineering(Upadhyay et al., 2025).

The table below summarizes the key characteristics, advantages, and limitations associated with each stage of evolution.

Approach	Key Characteristics	Techniques	Advantages	Limitations
Manual Inspection	Human visual assessment	Field observation, Expert diagnosis	Low cost (initial), Direct observation	Time-consuming, Labor-intensive, Error-prone, Subjective, Ineffective for early intervention
Traditional Image Processing	Rule-based, Handcrafted feature extraction	Canny edge detection, Histogram analysis, Thresholding	Non-destructive, Faster than manual	Struggles with complex backgrounds, Environmental variability, Limited generalization
Classical Machine Learning	Pattern recognition, Statistical models, Feature-dependent	SVM, Naïve Bayes, Decision Trees, KNN, Random Forest	Improved accuracy over traditional IP, Automates diagnosis	Requires handcrafted feature engineering, Scalability issues, Generalization challenges
Deep Learning	Automatic feature learning, Hierarchical representations	CNNs, Vision Transformers, Hybrid Models	High accuracy, Scalable, Real-time solutions, Handles large datasets	Data dependency, Computational complexity, Interpretability

Table 2: Summary of characteristics, advantages & limitations of different plant disease detection techniques

2.3 Image Processing and Computer Vision in Precision Agriculture

Recent developments in Image processing and computer vision have attracted numerous scientific studies in both crop health monitoring and disease detection, thereby improving precision agriculture. Remote sensing and image analysis techniques have demonstrated great promise in this domain, according to Zhang et al. (2019) and Shanmugam et al. (2017). This method involves the use of RGB or multispectral cameras to capture images across specific bands, such as near-infrared which is essential for crop analysis and calculation of Normalised Difference Vegetation Index (NDVI) to offer insights into plant vigour. Early foundational research by (Shanmugam et al., 2017) successfully employed canny edge detection algorithms and histogram analysis to identify crop diseases and trigger alert messages to agriculturalists, demonstrating the potential of image-processing techniques in precision agriculture.

2.4 Convolutional Neural Networks (CNNs) for Banana Disease Detection

Deep learning models such as Convolutional Neural Networks (CNNs) have emerged as a cornerstone of computer vision and image classification tasks. This is because they can automatically learn hierarchical representations directly from row images, which makes them effective for complex tasks such as object detection and image classification (Upadhyay et al., 2025)

CNNs are comprised of convolutional layers, which apply trainable filters to extract spatial features, pooling layers for reducing spatial dimensions and introducing translation invariance and fully connected/dense layers which are responsible for the final classification based on extracted features (Tugrul et al., 2022). With this kind of architecture, CNNs are capable of processing large image datasets in addition to withstanding variations in input, hence making them suitable for diverse agricultural imaging conditions.

A number of studies have consistently demonstrated the effectiveness of CNN models in banana disease detection. For example, (Narayanan et al., 2022) developed a hybrid CNN model that achieved 99% accuracy in classifying Black Sigatoka, bacterial wilt, and healthy banana leaves. (Selvaraj et al., 2019) employed transfer learning with pre-trained deep CNN models, achieving 92% accuracy for a broader range of banana leaf diseases. These results show how well Convolutional Neural Networks can be used to customise solutions for tasks involving the classification of pest and disease damage to banana plants.

2.5 Transfer Learning with Pretrained CNNs

Transfer learning (TL) approaches have been widely adopted in image classification tasks such as disease detection, mainly to address the challenge of limited labelled datasets. It involves leveraging knowledge from models pretrained on vast general image datasets like ImageNet and then fine-tuning them on a smaller domain-specific dataset, thereby reducing the computational costs and training time (Sambana et al., 2025)

Numerous studies have consistently shown that fine-tuned pre-trained models have always outperformed CNNs trained from scratch. For example, (Maeda-Gutiérrez et al., 2025) demonstrated that InceptionV3 has superior performance with 99.85% accuracy in Taro disease identification. Similarly, (Preotee et al., 2025) employed other transfer learning models like VGG19 and Xception, achieving impressive accuracies of up to 98.90% and 98.66% respectively.

The high performance of transfer learning models, such as InceptionV3, as compared to custom CNN models, is due to their ability to leverage pre-trained feature representations from vast image datasets like ImageNet. For practical agricultural applications like banana disease detection, where collecting large, annotated datasets is a challenge, Transfer learning can be the go-to approach.

2.6 Vision Transformers (ViTs) in Agricultural Imaging

With Vision Transformers (ViTs) there has been a paradigm shift in computer vision. ViTs adopt the highly successful self-attention mechanisms from Natural Language Processing (NLP) to image data processing. Unlike traditional deep learning models like CNNs, which rely on inductive biases such as local Receptive fields and pooling layers, which in most cases limit their performance in tasks where different assumptions hold. ViTs process images by converting them into sequences of fixed-sized patches/tokens. With this approach, ViTs are able to capture global contextual information and long-range dependencies across the entire image, hence enabling them to holistically understand visual patterns.

Different studies highlight the advantages of ViTs in agricultural imaging. For example, the Precision Leaf Analysis with Vision Transformers (PLA-ViT) framework demonstrated superior performance over CCN-based models in terms of accuracy (98.7% Vs 88.2%), disease localisation, inference time and computational efficiency (Murugavalli & Gopi, 2025).

2.7 Gray Level Co-occurrence Matrix (GLCM) for Texture Analysis

Gray Level Co-occurrence Matrix (GLCM) is a widely used image processing technique for texture analysis. GLCM quantifies image texture by analysing the spatial relationship between pixels. This is achieved by counting how often different grey-level pairs occur in a defined direction and the distance within an image, hence providing a numerical representation of the texture (Kumar & Singh, 2024). By combining GLCM with deep learning models, we can leverage an enhanced texture analysis capability, particularly effective in detecting diseases that manifest as changes in surface texture (Pang, Guo et al. 2019), such as Black Sigatoka. This integrated approach enables more accurate identification and characterisation of texture-related patterns, leading to improved disease diagnosis and monitoring.

Several studies have suggested that deep learning models tend to show significant improvement when fed with GLCM features. For example, (Sridhara Acharya, 2024) a hybrid CNN-SVM model that employed GLCM and Grey-level Difference Method (GLDM) features, achieved a 90% accuracy in crop disease classification.

The table below outlines the common GLCM texture features and their relevance in analysing crop/plant diseases

GLCM Feature	Description	Relevance to Plant Disease Detection
Contrast	Measures the intensity contrast between a pixel and its neighbor over the whole image. High values indicate large differences in intensity.	High contrast often indicates distinct lesion boundaries or abrupt changes from healthy to diseased tissue. Useful for identifying sharp edges of spots.
Dissimilarity	A linear measure of local intensity variations. Similar to contrast but less sensitive to extreme differences.	Reflects the variability of gray levels, useful for detecting irregular textures in diseased areas.
Homogeneity	Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. High values indicate a more uniform texture.	Low homogeneity might indicate irregular, mottled textures typical of certain fungal infections like Black Sigatoka, differentiating from smoother healthy tissue.
Energy (ASM)	Measures the sum of squared elements in the GLCM. High values indicate a more orderly, uniform texture.	Lower energy values can indicate a more disordered or complex texture, often associated with the progression of disease symptoms.
Correlation	Measures the linear dependency of gray levels in the co-occurrence matrix. High values indicate a strong linear relationship.	Reduced correlation can suggest a breakdown in the regular patterns of healthy tissue, indicative of disease.

Table 3: Relevance of common GLCM texture features

2.8 Hybrid Deep Learning Models for Enhanced Detection

Hybrid deep learning models are a result of combining the strengths of different architectures to overcome individual model limitations, thereby achieving superior overall performance. In studies where hybrid deep learning models have been employed, they have consistently shown high accuracies and improved robustness compared to standalone architectures. For example, MobilePlantViT with 99% accuracy (Tonmoy et al., 2025), ViT-B16 with 83.3% average accuracy (De Silva & Brown, 2023), and CNN-CBAM-SVM with 97.2% accuracy (Altalak et al., 2022). This consistent performance of hybrid models strongly indicates an underlying trend towards multi-modal and multi-architectural approaches as the future of advanced plant disease detection methods

2.9 Policy Framework and Agricultural Development Goals

Uganda's National Development Plan (NDP) III - Agro-industrialization Programme Implementation Action Plan (AgroPIAP) stresses the significance of data-driven decision-making in agriculture, aligning with national policies and strategic objectives. By using Convolutional Neural Networks (CNNs) for accurate banana disease detection, this research project seeks to support data driven agriculture and align with the national agricultural development agenda.

2.10 Gaps in Existing Research

While significant advancements have been made when using deep learning models for plant disease detection, several research gaps and practical challenges remain, for example, developing a model specifically tailored for banana diseases in Central Uganda. A region-specific model that considers the region's particular disease prevalence and environmental conditions is necessary to ensure precise disease detection and efficient management.

3.0. Research Methodology

This study implements a hybrid deep learning model that integrates Convolutional Neural Networks (CNNs), Vision Transformers (ViT), and Gray Level Co-occurrence Matrix (GLCM) texture analysis to improve the early detection of Fusarium Wilt and Black Sigatoka in smallholder farms in Central Uganda. The approach stresses interpretability of disease-specific features, resilience to limited data, and fit for resource-constrained implementation. The process from data collection to model deployment is summed up in Figure 1 below.

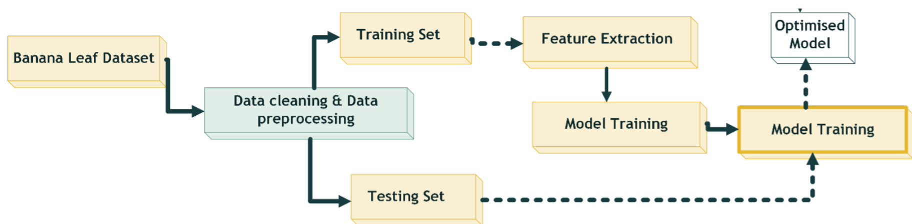


Figure 1: A Summary of Project Architecture

Data Collection and Preprocessing

The dataset used in this project consists of 17,068 labelled images of banana plants collected as part of the Lacuna Banana project (Mduma & Leo, 2023) by researchers and students from the Nelson Mandela African Institution of Science & Technology and the Tanzania Agricultural Research Institute (TARI). The images were captured using mobile phones equipped with the AdSurv application and focused on diagnosing two major banana diseases: Black Sigatoka (6,147 images) and Fusarium Wilt Race 1 (5,038 images), alongside healthy plants (5,883 images).

The preprocessing pipeline involved several steps to prepare the data for model training. First, the Images were resized to 256×256 pixels for CNN and InceptionV3 models and 224×224 pixels for ViT models to align with their respective architectural requirements. Pixel values were normalised to the range [0,1] using Min-Max normalisation to accelerate gradient descent during training. To further enhance dataset diversity and simulate field conditions, on-the-fly augmentation techniques were applied using TensorFlow's ImageDataGenerator. These included geometric transformations such as rotation (±30°), horizontal and vertical flipping (p=0.5), zooming (0.2 range), and shearing (0.1 rad), as well as photometric adjustments like brightness (±0.2) and HSV saturation (±0.1).

GLCM-Based Texture Feature Extraction

To augment deep learning features with texture cues, Gray Level Co-occurrence Matrix (GLCM) features were extracted from grayscale-converted Regions of

Interest (ROIs). Five texture features—contrast, dissimilarity, homogeneity, energy, and correlation were calculated and averaged to capture spatial dependencies in pixel intensities. After concatenating these GLCM features with flattened CNN or ViT features, the model was able to jointly learn textural and spatial patterns. The purpose of this hybrid approach was to increase the model's capacity to identify illness symptoms and boost classification precision.

Model Development

Three distinct approaches were explored for building the banana disease classifier, that is to say, a custom CNN architecture, transfer learning with InceptionV3, and a Vision Transformer (ViT).

3.3.1 Convolutional Neural Networks (CNN)

The custom CNN architecture was designed for lightweight deployment, making it suitable for resource-constrained environments. It consisted of three convolutional blocks, each with 3×3 kernels, ReLU activation, and L2 regularisation (=0.001). The first block used 32 filters followed by max-pooling (2×2), the second block used 64 filters with a dropout rate of 0.3, and the third block used 128 filters with global average pooling. The classification head comprised a dense layer with 128 units and ReLU activation, followed by a dropout layer (0.5) to prevent overfitting and a final dense layer with softmax activation for multiclass classification.

3.3.2 Transfer Learning

For this study, the InceptionV3 architecture was trained through a transfer learning approach. InceptionV3 is a deep convolutional neural network pre-trained on the large-scale ImageNet dataset. Transfer learning allows the model to leverage pre-learnt weights from ImageNet, which already capture rich representations of edges, textures, and object parts, and adapt them to the banana leaf disease classification task.

In this study, the final fully connected layers of InceptionV3 were replaced with custom dense layers tailored for three classes, i.e., Black Sigatoka, Fusarium Wilt, and Healthy. The convolutional base of the pre-trained model was retained to provide strong general feature extraction, while the new dense layers were trained on the banana leaf dataset. Fine-tuning was applied to the top layers of InceptionV3 to improve generalisation while avoiding overfitting.

This approach reduced training time and computational cost, while maintaining high accuracy by reusing the robust feature extraction capabilities of InceptionV3.

3.3.3 Vision Transformers (ViT)

The Vision Transformer (ViT) was adapted for this study; this model treats images as sequences of patches/tokens. Each image was split into 16×16 patches, which were projected into a 768-dimensional space using a trainable linear layer. A transformer encoder with six layers and eight attention heads processed these patch sequences/tokens. This approach enabled the model to capture complex patterns and relationships within the images, leveraging the strengths of transformer architectures.

Model Training

All models were trained in a GPU-enabled environment using Google Colab's parallel processing to speed up calculations in order to guarantee effective training. By consistently using the same training parameters, a fair and impartial comparison between models was preserved.

That is to say, using the Adam optimiser with a learning rate of 0.001, batch size of 32 samples, and training for 20 epochs. Sparse categorical cross-entropy was employed as the loss function for multiclass classification and implemented early stopping with a wait of 5 epochs to prevent overfitting, allowing the model to terminate training when no further improvements were observed. This standardised approach ensured a levelled ground for evaluating the performance of each mode.

Model Evaluation

All the models were evaluated using multiple performance metrics to assess their ability to classify banana diseases accurately. The overall correctness of the model predictions was examined through accuracy, and the model's performance in distinguishing between different banana diseases was analysed by computing precision, recall, and F1-score for each class. A confusion matrix was also generated to visually represent the model's performance across all categories, providing a detailed insight into its strengths and weaknesses in classifying various banana diseases.

4.0. Results and discussion

4.1. Convolutional Neural Network (CNN) Model Results

The classification report below highlights how the CNN model performed across the three different classes.

4.1.1 Classification Report

Table 4: CNN classification report

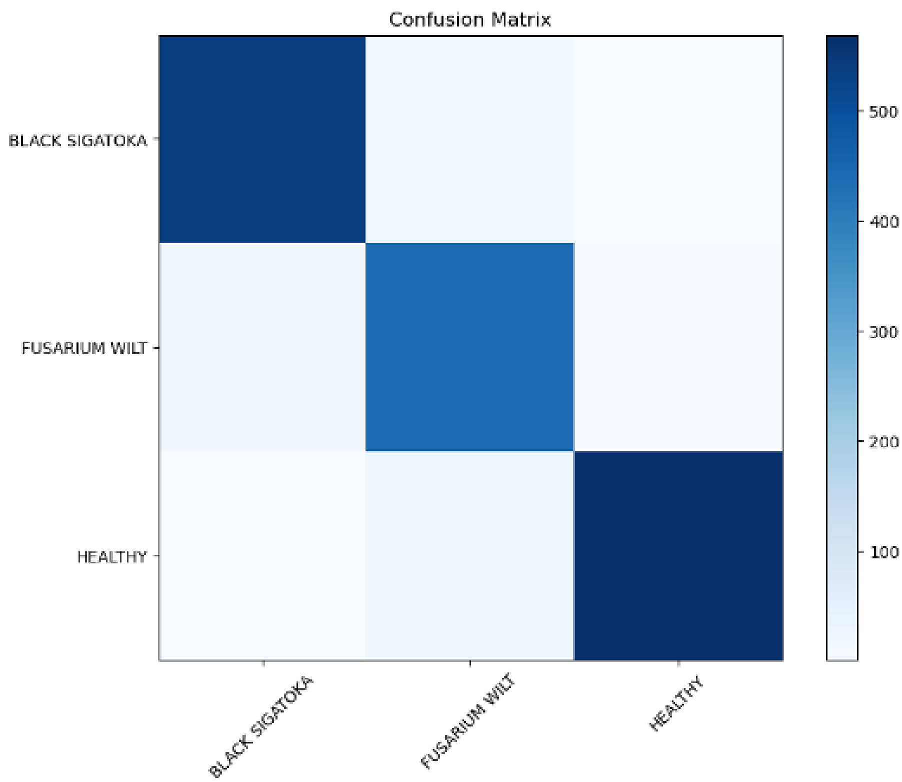
	precision	**recall**	**f1-score**	**support**
BLACK SIGATOKA	0.96	0.97	0.96	553
FUSARIUM WILT	0.92	0.94	0.93	466
HEALTHY	0.99	0.96	0.98	593
accuracy			0.96	1612
macro avg	0.96	0.96	0.96	1612
weighted avg	0.96	0.96	0.96	1612

The CNN model achieved an accuracy of 96%, showing strong but slightly uneven performance across the three disease categories. It performed exceptionally well in classifying Healthy banana leaves, attaining a precision of 0.99 and a recall of 0.96, resulting in an F1-score of 0.98. This indicates that the model accurately identified healthy plants with minimal false positives. However, its performance was relatively weaker in detecting Fusarium Wilt, where it recorded a precision of 0.92, a recall of 0.94, and an F1-score of 0.93. This suggests a higher rate of misclassification for Fusarium Wilt, potentially due to visual similarities with Black Sigatoka symptoms. This is also clearly visible in the confusion matrix below.

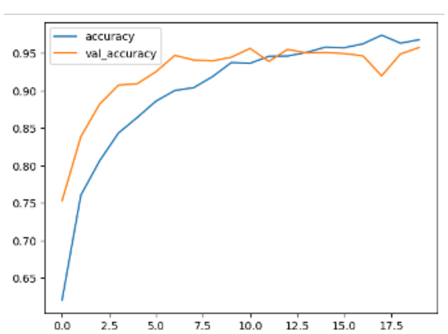
The model was reliable in classifying Black Sigatoka, with metrics around 0.96–0.97. Overall, the macro and weighted averages were consistent at 0.96, indicating balanced performance across the dataset, albeit with room for improvement in disease-specific classification, especially for Fusarium Wilt

4.2 Transfer Learning Model Results

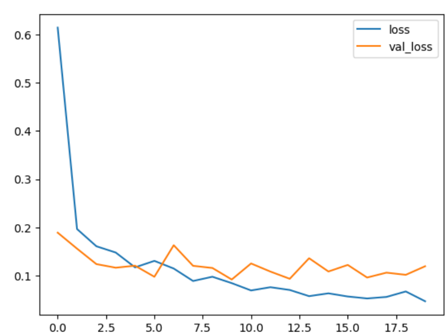
Using a transfer learning approach with a pre-trained **InceptionV3 model** improved the model's classification performance across all classes, as shown in the classification report below.



(a) Confusion matrix



(b) CNN training & validation accuracy



(c) CNN training & validation loss

Figure 2: CNN Model performance on the banana-leaf disease dataset.

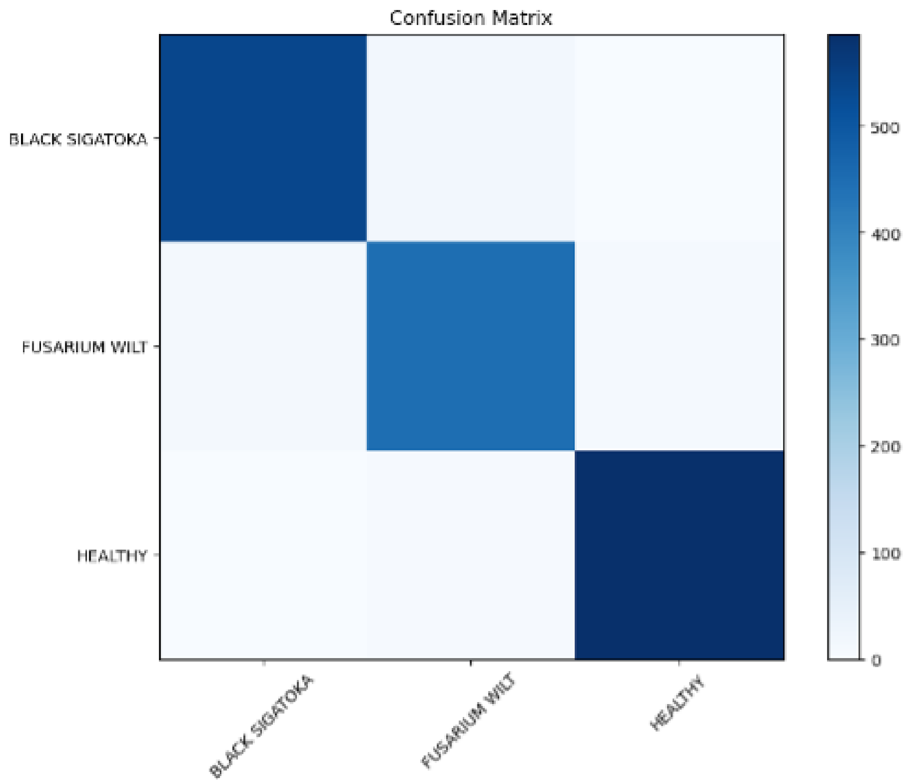
4.2.1 Classification Report

Table 5: InceptionV3 classification report

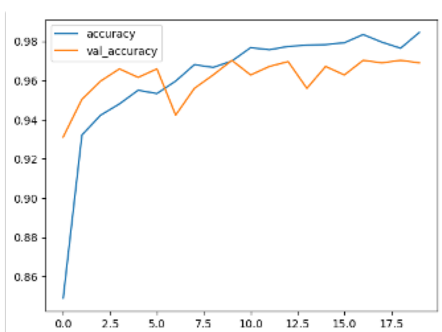
	precision	recall	f1-score	support
BLACK SIGATOKA	0.98	0.97	0.97	553
FUSARIUM WILT	0.96	0.96	0.96	466
HEALTHY	0.98	0.99	0.99	593
accuracy			0.97	1612
macro avg	0.97	0.97	0.97	1612
weighted avg	0.97	0.97	0.97	1612

The model based on Transfer Learning yielded an improved overall accuracy of 97%, outperforming the CNN across all three categories. It demonstrated powerful performance on Healthy samples, with both precision and recall reaching 0.98–0.99, translating to a near-perfect F1 score of 0.99. This indicates the model was both accurate and consistent in identifying healthy leaves. The classification of Black Sigatoka was also excellent, with a precision of 0.98 and a recall of 0.97, resulting in a robust F1-score of 97%.

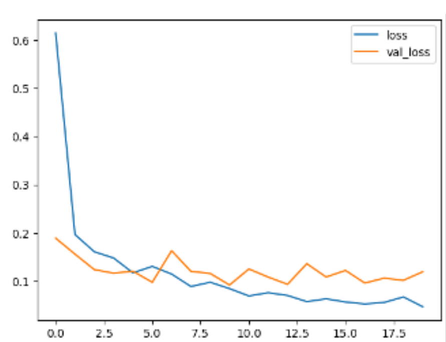
The CNN's shortcomings were addressed by this model, which showed a balanced improvement for Fusarium Wilt with a precision and recall of 0.96. The macro and weighted averages of 0.97 show that all classes performed consistently. The Transfer Learning model significantly improved accuracy and reliability, most likely due to the pre-learned feature representations that improved generalisation.



(a) Confusion matrix



(b) InceptionV3 training & validation accuracy



(c) InceptionV3 training & validation loss

Figure 3: InceptionV3 Model performance on the banana-leaf disease dataset.

4.3 Vision Transformer (ViT) Model Results

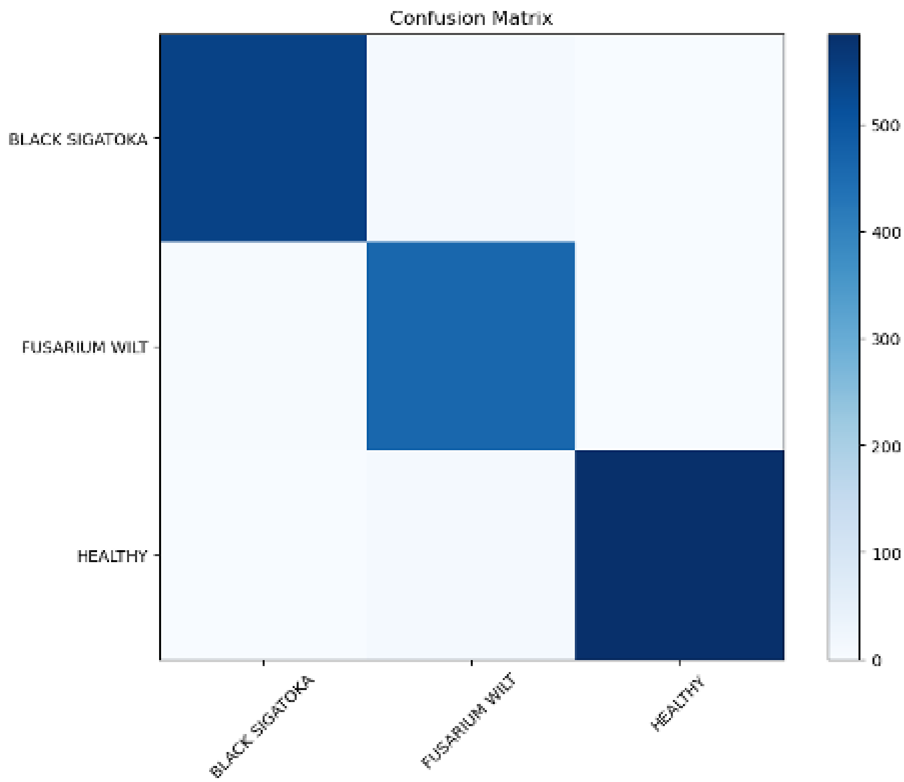
The ViT model outperformed both the CNN and InceptionV3 models, achieving a remarkable overall accuracy of 99%. Its performance was also consistently high across all classes, as shown in the classification report below

4.3.1 Classification Report

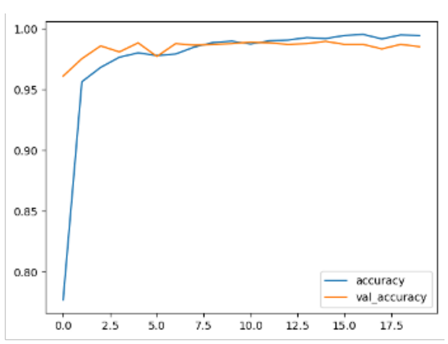
Table 6: ViT classification report

	precision	**recall**	**f1-score**	**support**
BLACK SIGATOKA	0.99	0.98	0.99	553
FUSARIUM WILT	0.96	0.99	0.98	466
HEALTHY	0.99	0.99	0.99	593
accuracy			0.99	1612
macro avg	0.98	0.99	0.98	1612
weighted avg	0.99	0.99	0.99	1612

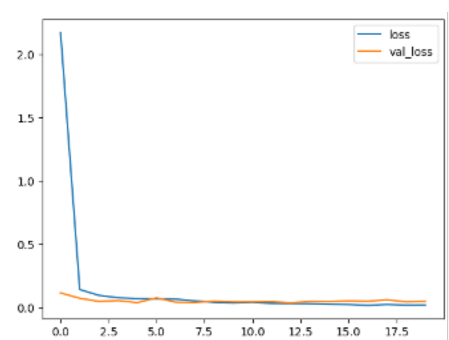
For Healthy leaves, the model recorded near-perfect metrics with precision and recall at 0.99 and a matching F1 score, indicating very few classification errors. In identifying Black Sigatoka, the model maintained high precision (0.99) and recall (0.98), again showing robust reliability. With an F1-score of 0.98 and a recall of 0.99, the ViT model remarkably outperformed other models in classifying Fusarium Wilt, demonstrating a strong ability to identify this frequently misdiagnosed disease. In addition to confirming excellent balance, the weighted and macro averages of 0.98 to 0.99 highlight the ViT model's superior generalisation ability.



(a) Confusion matrix



(b) ViT training & validation accuracy



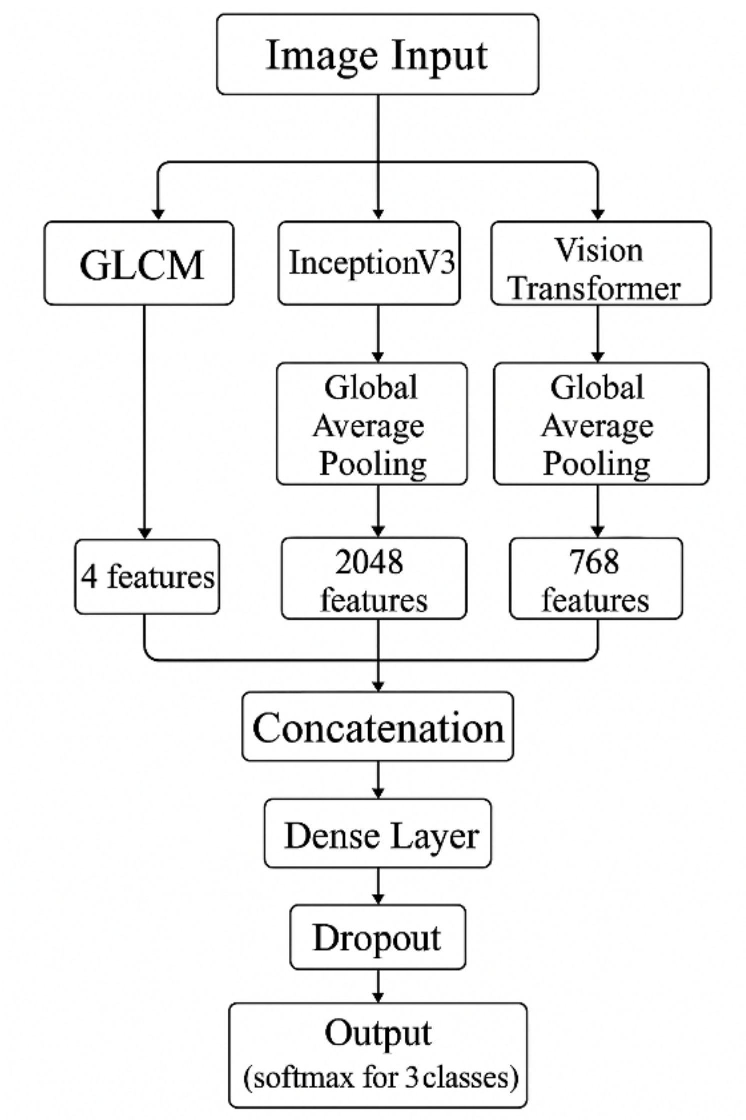
(c) ViT training & validation loss

Figure 4: ViT model performance on the banana-leaf disease dataset.

5. Conclusion

In conclusion, this research proposes a hybrid model that integrates the computational efficiency of the InceptionV3 architecture with the Vision Transformer's capability to represent images as sequences of fixed-size patches, enabling local feature extraction and global contextual understanding.

5.1 Hybrid Model Conceptual Diagram



The diagram above presents a hybrid deep-learning architecture for banana disease image classification. The model begins with an Image Input, which is simultaneously fed into three parallel processing paths. The first path applies Gray-Level Co-occurrence Matrix (GLCM) techniques to extract texture features, which are helpful in identifying tiny differences between disease classes.

The second path processes the image using InceptionV3, a convolutional neural network known for its computational efficiency. After feature extraction, Global Average Pooling is applied to reduce the spatial dimensions and produce a compact 2048-dimensional feature vector. The third path utilises the Vision Transformer (ViT), which interprets the image as a sequence of fixed-size patches and uses self-attention mechanisms to model global relationships across the image. The ViT output also undergoes Global Average Pooling, resulting in a 768-dimensional feature vector.

The outputs from GLCM (4 features), InceptionV3 (2048 features), and ViT (768 features) are then merged through a Concatenation layer, forming a unified feature vector that captures both texture and deep semantic information. This concatenated vector is passed through a Dense layer, which enables the network to learn complex, non-linear interactions between features. To prevent overfitting, a Dropout layer is applied before the final prediction is made through a Softmax Output layer, which classifies the image into one of three categories—Healthy, Fusarium Wilt, or Black Sigatoka.

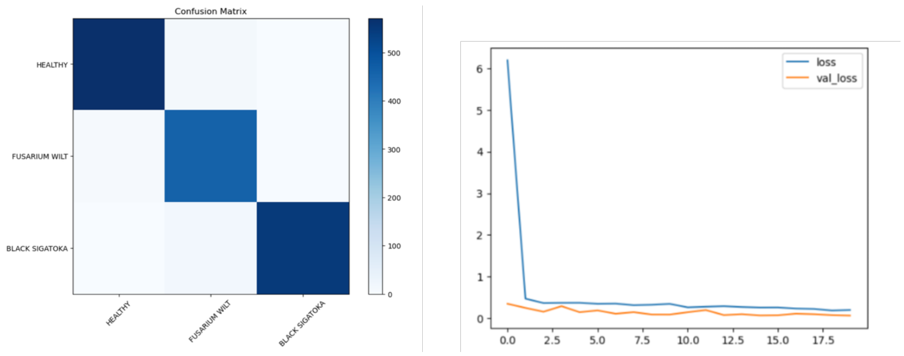
This hybrid approach exploits the strengths of both traditional feature engineering and deep learning to improve classification accuracy and generalisation.

5.2 Classification Report for the Hybrid Model

The results from the hybrid model, as shown in the classification report below, indicate strong and balanced performance across all three categories—Healthy, Fusarium Wilt, and Black Sigatoka.

Table 7: Hybrid model classification report

	precision	**recall**	**f1-score**	**support**
BLACK SIGATOKA	0.99	0.97	0.98	553
FUSARIUM WILT	0.95	0.98	0.97	466
HEALTHY	0.99	0.98	0.99	593
accuracy			0.98	1612
macro avg	0.98	0.98	0.98	1612
weighted avg	0.98	0.98	0.98	1612



(a) Hybrid Model Confusion Matrix

(b) Hybrid model training & validation loss

Figure 5: Hybrid model performance on the banana-leaf disease dataset.

References

Altalak, M., Uddin, M. A., Alajmi, A., & Rizg, A. (2022). A Hybrid Approach for the Detection and Classification of Tomato Leaf Diseases. *Applied Sciences (Switzerland)*, *12*(16). https://doi.org/10.3390/app12168182

Damale, P., Panchal, S., Gawand, P., Waghmare Guide, P., & Deshmukh, S. (2022). Crop Yield Prediction and Disease Detection Using Machine Learning. *JETIR*. www.jetir.orgh620

De Silva, M., & Brown, D. (2023). Multispectral Plant Disease Detection with Vision Transformer-Convolutional Neural Network Hybrid Approaches. *Sensors (Basel, Switzerland)*, *23*(20). https://doi.org/10.3390/s23208531

Domingues, T., Brandão, T., & Ferreira, J. C. (2022). Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. In *Agriculture (Switzerland)* (*12*(9)). https://doi.org/10.3390/agriculture12091350

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. http://arxiv.org/abs/2010.11929

FAO. (2019). The PlanTVillage PlaTform. www.fao.org/food-chain-crisis

Hassan, S. M., & Maji, A. K. (2022). Plant Disease Identification Using a Novel Convolutional Neural Network. *IEEE Access*, *10*, 5390–5401. https://doi.org/10.1109/ACCESS.2022.3141371

International Trade Administration U.S. Department of Commerce. (2023, October 13). https://www.trade.gov/country-commercial-guides/uganda-agricultural-sector.

Kumar, R., & Singh, V. (2024a). Combining GLCM Features with SVM Classification for Improved Accuracy in Cotton Crop Disease Detection. *Proceedings - 2024 2nd International Conference on Advanced Computing and Communication Technologies (ICACCTech 2024)*, 110–115. https://doi.org/10.1109/ICACCTech65084.2024.00028

Kumar, R., & Singh, V. (2024b). Combining GLCM Features with SVM Classification for Improved Accuracy in Cotton Crop Disease Detection. https://doi.org/10.1109/ICACCTech65084.2024.00028

Maeda-Gutiérrez, V., Oropeza-Valdez, J. J., Reveles-Gómez, L. C., Padron-Manrique, C., Resendis-Antonio, O., Solís-Sánchez, L. O., Guerrero-Osuna, H. A., & Olvera Olvera, C. A. (2025a). AI-Driven Plant Health Assessment: A Comparative Analysis of Inception V3, ResNet-50 and ViT with SHAP for Accurate Disease Identification in Taro. *Agronomy*, *15*(1). https://doi.org/10.3390/agronomy15010077

Maeda-Gutiérrez, V., Oropeza-Valdez, J. J., Reveles-Gómez, L. C., Padron-Manrique, C., Resendis-Antonio, O., Solís-Sánchez, L. O., Guerrero-Osuna, H. A., & Olvera Olvera, C. A. (2025b). AI-Driven Plant Health Assessment: A Comparative Analysis of Inception V3, ResNet-50 and ViT with SHAP for Accurate Disease Identification in Taro. *Agronomy*, *15*(1). https://doi.org/10.3390/agronomy15010077

Marin, D. H., Romero, R. A., Guzmán, M., & Sutton, T. B. (2003). Black Sigatoka: an increasing threat to banana cultivation. *Plant Disease*, *87*(3), 208–222.

Mduma, N., Laizer, H., Loyani, L., Macheli, M., Msengi, Z., Karama, A., Msaki, I., Sanga, S., Jomanga, K., & Judith, L. (2022). The Nelson Mandela African Institution of Science and Technology Bananas dataset. *Harvard Dataverse*. https://doi.org/10.7910/DVN/LQUWXW

Mduma, N., & Leo, J. (2023). Dataset of banana leaves and stem images for object detection, classification and segmentation: A case of Tanzania. *Data in Brief*, *49*. https://doi.org/10.1016/j.dib.2023.109322

Mehdipour, S., Mirroshandel, S. A., & Tabatabaei, S. A. (2025). Vision Transformers in Precision Agriculture: A Comprehensive Survey. http://arxiv.org/abs/2504.21706

Murugavalli, S., & Gopi, R. (2025). Plant leaf disease detection using vision transformers for precision agriculture. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-05102-0

Narayanan, K. L., Krishnan, R. S., Robinson, Y. H., Julie, E. G., Vimal, S., Saravanan, V., & Kaliappan, M. (2022). Banana Plant Disease Classification Using Hybrid Convolutional Neural Network. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/9153699

National Planning Authority (NPA). (2020). NDPIII AGRO-INDUSTRIALIZATION PROGRAMME IMPLEMENTATION ACTION PLAN.

PARM. (2017). Platform for Agricultural Risk Management Managing risks to improve farmers' livelihoods Crop pests and disease management in Uganda: status and investment needs Full Report.

Pegg, K. G., & Langdon, P. W. (1987). Fusarium wilt (Panama disease): a review. *Banana and Plantain Breeding Strategies*, 119–123.

Preotee, F. F., Sarker, S., Refat, S. R., Muhammad, T., & Islam, S. (2025). An Approach Towards Identifying Bangladeshi Leaf Diseases through Transfer Learning and XAI. https://doi.org/10.1109/ICCIT64611.2024.11022335

Rietveld, A. M., Jogo, W., Mpiira, S., & Staver, C. P. (2014). The effect of Banana Xanthomonas Wilt on beer-banana value chains in central Uganda: An exploratory study. https://api.semanticscholar.org/CorpusID:140658750

Salman, Z., Muhammad, A., & Han, D. (2025). Plant disease classification in the wild using vision transformers and mixture of experts. *Frontiers in Plant Science*, *16*. https://doi.org/10.3389/fpls.2025.1522985

Sambana, B., Nnadi, H. S., Wajid, M. A., Fidelia, N. O., Camacho-Zuñiga, C., Ajuzie, H. D., & Onyema, E. M. (2025). An efficient plant disease detection using transfer learning approach. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-02271-w

Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., & Blomme, G. (2019). AI-powered banana diseases and pest detection. *Plant Methods*, *15*(1). https://doi.org/10.1186/s13007-019-0475-z

Shanmugam, L., Adline, A., Aishwarya, N., & Krithika, G. (2017). *Disease detection in crops using remote sensing images.* https://doi.org/10.1109/TIAR.2017.8273696

Sherony, K. R., Knowles, G. J., & Boyd, R. (1991). The Economic Impact of Crop Losses: A Computable General Equilibrium Approach.

Singh, R., Agarwal, G., Sanghi, A., & Srivastava, R. P. (n.d.). PLANT DISEASE DETECTION USING HYBRID DEEP LEARNING APPROACH.

Sridhara Acharya, P. (2024). MobileNetV3Small-CM FusionNet: A Lightweight Deep Learning Framework for Multi-Class Arecanut Disease Classification Using Feature Fusion. In *Journal of Information Systems Engineering and Management* (*2025*(4)). https://www.jisem-journal.com/

Tonmoy, M. R., Hossain, Md. M., Dey, N., & Mridha, M. F. (2025). MobilePlantViT: A Mobile-friendly Hybrid ViT for Generalized Plant Disease Image Classification. http://arxiv.org/abs/2503.16628

Tripathi, L., Mwangi, M., Abele, S., Aritua, V., Tushemereirwe, W. K., & Bandyopadhyay, R. (2009). Xanthomonas wilt: a threat to banana production in East and Central Africa. *Plant Disease*, *93*(5), 440–451.

Tugrul, B., Elfatimi, E., & Eryigit, R. (2022). Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. *Agriculture*, *12*(8). https://doi.org/10.3390/agriculture12081192

Upadhyay, A., Chandel, N., Singh, K., Chakraborty, S., Nandede, B., Kumar, M., Subeesh, A., Upendar, K., Salem, A., & Elbeltagi, A. (2025). Deep learning and computer vision in plant disease detection: a comprehensive review of techniques, models, and trends in precision agriculture. *Artificial Intelligence Review*, *58*. https://doi.org/10.1007/s10462-024-11100-x

V, B. S. (2012). Grey Level Co-Occurrence Matrices: Generalisation and Some New Features. *International Journal of Computer Science, Engineering and Information Technology*, *2*(2), 151–157. https://doi.org/10.5121/ijcseit.2012.2213

Xiao, L., Tiancong, Z., Jia, Y., Nie, X., Wang, M., & Shao, X. (2025). YOLO-RS: Remote Sensing Enhanced Crop Detection Methods. http://arxiv.org/abs/2504.11165

Yogeshwari, M., & Thailambal, G. (2023). Automatic feature extraction and detection of plant leaf disease using GLCM features and convolutional neural networks. *Materials Today: Proceedings*, *81*, 530–536.

Zhang, J., Huang, Y., Pu, R., Gonzalez-Moreno, P., Yuan, L., Wu, K., & Huang, W. (2019). Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, *165*, 104943. https://doi.org/https://doi.org/10.1016/j.compag.2019.104943