OPEN ACCESS
International Journal of Innovations in Science & Technology

Deep Transformer Based Anomaly Detection in Sugarcane Red Rot Disease Using Spectral Leaf Images

Hamza Haider[1], Rabia Tehseen[1], Muhammad Chaudhry[1], Uzma Omer[2]
[1]Department of Computer Science, University of Central Punjab, Lahore, Pakistan
[2]Department of Information Sciences, University of Education, Lahore, Pakistan
*Correspondence: rabia.tehseen@ucp.edu.pk

Citation | Haider. H, Tehseen. R, Chaudhry. M, Omer. U, "Deep Transformer Based Anomaly Detection in Sugarcane Red Rot Disease Using Spectral Leaf Images", IJIST, Vol. 8 Issue. 3 pp 1001-1015, June 2026

Received | April 09, 2026 **Revised |** May 12, 2026 **Accepted |** May 19, 2026 **Published |** June 02, 2026.

This study presents a hyperspectral framework using transformers for the detection of Red Rot disease in sugarcane through the use of 31-band hyperspectral leaf images. A set of 400 hyperspectral images was used, with 200 healthy and 200 diseased sugarcane leaves, all within the visible region (400 to 700 nm) being used to develop the proposed method, which includes the use of MST++ spectral reconstruction, wavelength-aware positional encoding, spectral/spatial feature extraction and transformer-based classification for automated monitoring of crop diseases. The Hybrid model outperforms the Spectral-Spatial Transformer (SST) model in classification. Overall classification accuracy is 83.33% for the Hybrid model, while SST achieves 65%. The classified samples of sugarcane that are affected by Red Rot show improved performance on recall (0.90), F1-score (0.84), and sensitivity (86.67%) with the Hybrid framework, in comparison to reduced false negatives and false positives in the confusion matrix analyses. The SST model achieved a validation accuracy of approximately 70%, while the Hybrid models exhibited consistent behavior and evaluated close to 80% in terms of their validation accuracy, with lower training and validation loss values, implying that they performed better in the aspects of convergence and generalization. The Hybrid framework also showed an increase of nearly 18.33% in its classification accuracy when compared with the SST architecture. The proposed Hybrid framework represents a scalable, explainable, and intelligent approach to precision agriculture, monitoring crop diseases, and managing sugarcane disease in a sustainable manner.

Keywords: Hyperspectral Imaging, Red Rot Disease, Spectral–Spatial Learning, Attention Mechanism.

Introduction:

The importance of agriculture to food security and economic sustainability around the world is reflected in the fact that monitoring the condition of crops is now an important area of research in today's agricultural systems. With developments in AI, deep learning, and computer vision, there have been major advancements to automated methods for detecting crop diseases. Specifically, the use of hyperspectral imaging is a new and effective means of assessing the health of plants because it can measure the amount of light in a number of different parts of the spectrum (beyond the visible range captured by standard RGB images). Compared with standard image analysis methods, hyperspectral imaging provides the ability to detect small biochemical and structural changes in the cells of plants, making this method ideal for detecting diseases at early stages and making it valuable for precision agriculture applications.

Sugarcane disease detection has captured worldwide attention within the agriculture industry because of its large economic impact. One of the most destructive diseases on the sugarcane crop is Red Rot, which is a very serious disease and is responsible for significant yield losses and declines in sugarcane quality. Diagnosis of sugarcane disease is currently done through field inspection and later laboratory testing. This process is time-consuming and often fails to detect disease at early stages. Recent advancements in deep learning, hyperspectral imaging, transformer architectures, and spectral–spatial feature learning have opened the door to developing intelligent systems for automatic and highly accurate detection of symptoms associated with sugarcane disease.

The development of intelligent systems for crop disease detection has been aided by advancements in hyperspectral imaging and transformer-based methods. Researchers have shown SSTN and SS-TMNet to be highly effective at extracting spectral and spatial relationships from hyperspectral images [1][2][3]. Hybrid CNN and vision transformer approaches have also enhanced robustness and accuracy of crop disease classification [4][5]. Additionally, hybrid deep learning models have also been shown to be effective for detecting sugarcane diseases and monitoring precision agriculture [6][7]. Moreover, spectral-spatial attention mechanisms have enhanced model interpretability through identification of spectral discrepancies indicative of disease [1]. Nonetheless, limited research exists on transformer-based hyperspectral anomaly detection methods for early detection of Red Rot disease in sugarcane [8][7]. As a result, this work addresses the need for development of an interpretable transformer based hyperspectral anomaly detection framework that will support accurate early diagnosis of sugarcane disease and intelligent agricultural production monitoring within the precision agriculture domain.

Research Objectives:

The main objective of this study is to develop a smart and interpretable transformer-based anomaly detection framework for the early detection of Red Rot disease in sugarcane using hyperspectral imaging. The proposed framework aims to combine spectral–spatial feature learning, self-attention mechanisms, and anomaly detection techniques to accurately identify disease-related abnormalities before visible symptoms appear. The following objectives are proposed:

Develop a transformer-based model for classifying healthy and Red Rot affected sugarcane leaves.

Utilize multi-head self-attention mechanisms to extract spectral–spatial features from hyperspectral images.

Integrate an anomaly detection head to identify spectral abnormalities associated with Red Rot disease.

Research Gap and Novelty:

Significant progress has been made with the use of deep learning and hyperspectral imaging for plant disease detection; there are still many limitations in this area of research. For instance, a vast majority of current research uses classical CNN-based classification methods to extract local features from the images. Although these conventional methods work very well at extracting local feature information from images, they do not support the modelling of long-range spectral–spatial dependencies that exist in hyperspectral images. A number of current methods for plant disease detection (sugarcane) rely on the analysis of visible symptoms of disease, hence, did not detect the infection before visible symptoms are high in the disease life cycle (i.e. the early pre-symptomatic stage of the disease). Research on the use of transformer-based hyperspectral images is limited. The majority of these methods have been created to analyze general types of images. Thus, there is a limited number of studies focused specifically on the development of anomaly-driven transformer frameworks for the analysis of Red Rot disease in sugarcane. Furthermore, the lack of interpretability and explainability for AI technologies in agriculture is a major challenge and limits their usability for precision farming and decision making for agribusinesses.

Although advancements have been made in deep learning systems and hyperspectral imagery for detecting agricultural disease, there are still limitations with existing systems. Most existing systems rely on CNN-based approaches that mainly focus on extracting local features, which means that they do not capture long-range spectral–spatial dependencies that are often present in hyperspectral datasets [4][9]. Although transformer-based hyperspectral classifiers like SSTN and SS-TMNet have improved the ability of models to learn spectral–spatially, none of these models are capable of detecting Red Rot disease in sugarcane early on [2][3]. Furthermore, most agricultural disease detection systems do not have explainable or anomaly-aware learning approaches to identify spectral anomalies before visible symptoms occur. This paper proposes a new transformer-based framework for hyperspectral anomaly detection in order to accurately and interpretably detect Red Rot disease in sugarcane before the development of symptoms.

This research proposes a novel framework for early Red Rot detection in sugarcane plants. This is called the anomaly detection framework. The anomaly detection framework uses dual-attention transformer neural networks with images that were taken using hyperspectral imaging. With this system, farmers will be able to find sugarcane plants that are starting to develop red rot before they can see any visible signs of it. Using hyperspectral imaging along with transformer learning, anomaly detection, and explainable artificial intelligence will allow for scalable, intelligent methods for precision agriculture and sustainable crop management.

Contribution:

This study adds to the existing body of research focused on developing an accurate method for detecting Red Rot disease early in sugarcane crops through hyperspectral images by using a transformer-based imaging framework. The combination of MST++ reconstructed spectra and a wavelength-aware positional encoding method will allow for improved discrimination of healthy and diseased plants via 31 band hyperspectral images cubed together. In this research, two state-of-the-art deep learning models are applied that use spectral and spatial data, the Spectral-Spatial Transformer (SST) and Hybrid 3D CNN + Transformer methods, which were used together to evaluate the efficacy of each model for analyzing hyperspectral images. Furthermore, the attention-based learning and spectral features will improve the interpretability of the model and allow for scalable and reliable options within the context of intelligent crop disease monitoring systems (ICDMS) and precision agriculture (PA).

Literature Review:

Traditional machine-learning techniques are recognized as some of the first used to classify the plant pathogens associated with hyperspectral imaging data. Studies have shown that machine-learning algorithms utilize spectral data of agricultural products for their monitoring and surveying [1][2]. The major limitation of using these techniques is that they require handcrafted feature extraction which restricts the potential for capturing the complex spectral-spatial correlation associated with hyperspectral datasets [10][3].

Advanced hyperspectral instrumentation has allowed for the identification and detection of plant disease through detailed spectral analysis of the leaves of crops using hyperspectral imaging techniques and spectral-spatial attention mechanisms for the identification of diseases of sugar cane (e.g., smut and mosaic disease) [1][8]. In addition, hybrid methods combining spectral and transformer frameworks have demonstrated substantial improvements in the prediction of crop disease through enhanced feature representation and classification performance [4][5][6][7].

A number of studies conducted numerous types of analysis on both the biological and pathological characteristics of the disease known as Red Rot that affects sugarcane when it is infected with the fungus, Colletotrichum falcate. The variability of the pathogen, its ability to cause disease and how different sugarcane varieties responded to inoculation were analyzed through various methods [11][12][13]. In addition, the use of molecular diagnostic techniques, including Looped-Mediated Isothermal Amplification (LAMP), was proposed as a quick and sensitive way to identify the presence of this disease [14].

Automated systems for identifying diseases in plants have been greatly improved by the use of deep learning methods. Convolutional Neural Networks (CNN's) have shown a high degree of success in identifying the textures and visual features of diseases on agronomical images [9][15]. In comparative studies, deep learning models have been shown to outperform traditional machine learning methods for accurately identifying and classifying diseases and their robustness [16][17]. Hyperspectral imaging has also been successfully used along with deep learning to provide early indications of disease when monitoring crops [18][19].

Recently, transformers have demonstrated an impressive performance in the classification of hyperspectral images as they can model long range dependencies between pixels through self-attention. The original transformer architecture provided by Vaswani et al. has served as the basis for the majority of attention-based deep learning models [20]. In addition to this original work, new types of attention (such as Squeeze-and-Excitation Networks, and residual learning) have enhanced the ability to extract features from hyperspectral data for classification [21][22]. Spectral-Spatial Attention Networks focused on providing higher accuracies for hyperspectral data classification by extracting informative spectral bands and spatial regions from the images [3].

There are many types of advanced transformer-based hyperspectral classification models that have demonstrated superior spectral and spatial feature learning performance; for instance, Spectral-Spatial Transformer Networks (SSTN), SS-TMNet, and Multi-Range Spectral-Spatial transformers [3][2][10]. All of these networks utilize multi-range transformer architectures to extract global contextual information from the pixels in hyperspectral images, which increases their robustness to variations in the data. In addition, researchers have developed hybrid architectures from the Vision Transformer and convolutional neural networks to classify agricultural diseases by employing a combination of local feature extraction and transformer-based contextual learning [4][6].

Recent research has been done using Convolutional Neural Networks (CNNs) and Vision Transformers to detect diseases in sugarcane with an aim of improving both the classification accuracy and the speed at which early detection of diseases are made. The hybrid transformer-based approaches that have been developed for detecting several types of sugarcane disease by using only leaf image data sets have produced very high results [23]. While there has been a large amount of research done on CNNs and Vision Transformers there has been little focus on using a transformer based hyperspectral approach for early detection of Red Rot disease in sugarcane. This highlights the uniqueness and novelty of the proposed research framework.

Methodology:

This study presents an end-to-end approach for detecting Red Rot disease in sugarcane leaves using hyperspectral imagery. The proposed framework utilizes the complete process of hyperspectral image generation, spectral/spatial pre-processing, transformer-based deep learning, and anomaly-aware classification in order to provide binary classification (healthy/infected) of sugarcane samples. The proposed framework is designed to optimally capture both spectral and spatial information from the hyperspectral images and also enhance interpretability and classification accuracy for precision agriculture.

In this study, early detection refers to the ability to identify sugarcane samples that have been affected by Red Rot before any visible symptoms are evident by means of analyzing hyperspectral spectral variation rather than looking at the severity of visual symptoms. The dataset that was used in this work comprises two labelled classes (healthy leaves and Red Rot infected leaves) and does not have any other experimentally validated classes such as early infected, medium infected, or severely infected samples. As a result, only binary classification can be performed in the proposed framework between healthy and infected samples using 31-band hyperspectral images. A hyperspectral image consists of several images containing different spectral bands of light reflectance when they are examined at different wavelengths across the spectrum typically 400 nm to 700 nm in this study. Compared to RGB images, which only contain three channels of color, hyperspectral images can contain multiple spectral bands because they measure several different wavelengths of light in a continuous range, making them suitable for studying the spectral, biochemical and structural characteristics of plant tissues. The use of hyperspectral images makes them an ideal tool for the early detection of plant diseases and precision agriculture, thus allowing the proposed framework to be able to support the early detection of disease at the spectral level.

Figure 1 illustrates the full workflow of the proposed hyperspectral image classification framework for detecting Red Rot disease in sugarcane leaves, including the process starting with RGB images of sugarcane leaves and passing them through the MST++ spectral reconstruction module to create hyperspectral image cubes consisting of 31 spectral bands that correspond to the visible wavelength range. This workflow includes the creation of hyperspectral image cubes through spectral reconstruction, storage of the cubes in .mat file format, and performance of preprocessing operations on each hyperspectral cube. Centre patches of size 5×5×31 and 7×7×31 were extracted from each hyperspectral cube to maintain the local spectral/spatial information while decreasing the computational complexity of the process.

A wavelength-aware encoding will be incorporated after preprocessing the hyperspectral data, permitting the 31 spectral bands to represent their sequential relationship. This encoding allows transformers to learn the spatial placement of the different wavelengths, thus optimizing the learning of spectral features. In addition, the encoded hyperspectral patches will be provided to two different feature extraction models. Using self-attention, the SST model is capable of capturing long-range dependencies among spectral features. The Hybrid model is a combination of the 3D CNN for extracting local spectral-spatial features along with a lightweight transformer encoder (for learning global contextual relationships). Finally, the features will be processed through fully connected layers to classify either "healthy" or "red rot" leaves of sugarcane.

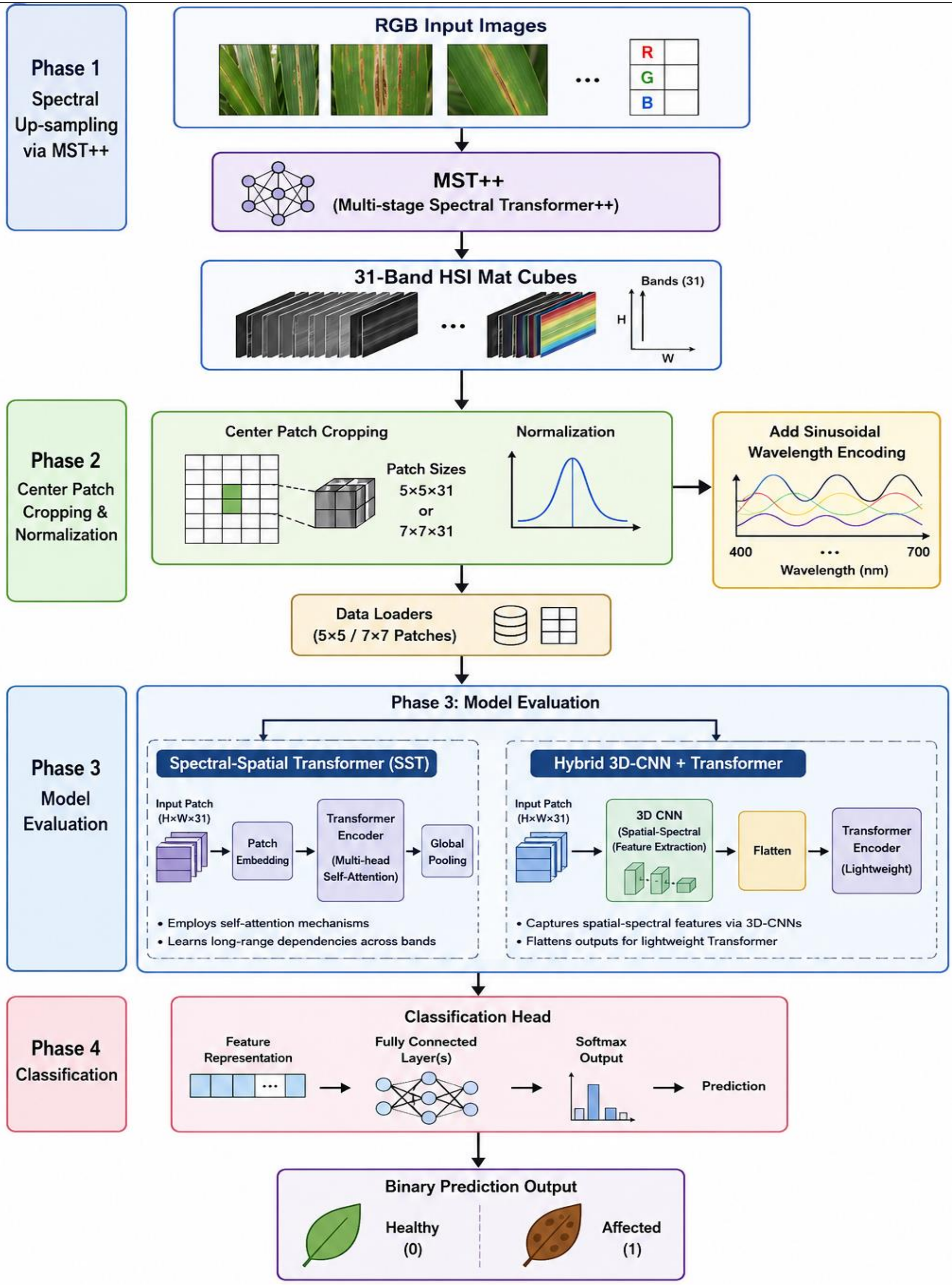

Figure 1. Workflow of proposed methodology

Data Preprocessing and Normalization:

The initial data preprocessing will consist of two main phases: converting the RGB image into an HSI representation (by applying spectral reconstruction) and preparing the input for training a prediction using the constructed HSI image. During this first phase, each RGB image will be scanned for duplicates, and all duplicate images will be removed to provide the model with clean input data (only unique images). After cleaning the input images, they will be reconstructed from RGB to HSI using the MST++ reconstruction software to produce one hyperspectral cube per input image using 31 channels within the visible range of wavelengths (400-700 nm). Based on the input parameters for each image during reconstruction, all reconstructed images will be stored in mat format as compressed files for additional processing prior to model training.

The second phase includes preparing the .mat file hyperspectral cubes for use with machine learning and deep learning algorithms. The normalization for all hyperspectral cubes is accomplished using peak value normalization, each matrix is divided by its maximum absolute intensity value, so its data values are within a bounded range from 0 to 1. After normalization, 70% of the dataset is used for training, while the remaining 30% is divided equally between validation and testing datasets by using stratified sampling to maintain a consistent proportion of healthy and diseased samples throughout all three datasets. Center patch cropping reduces the computing analysis time by reducing the size of the data and concentrating on local and significant patterns of disease within the original 256×256 hyperspectral image. The center patch cropping is accomplished using 5×5 pixel and 7×7-pixel patches with respect to the 31 spectral bands associated with each patch. Using a physics informed sinusoidal wavelength encoding to represent the spectral positions of the 31 bands provides additional information to transformer-based models so that they can learn the progression and relationships of the wavelengths as they are making classifications.

Spectral Up-Sampling Using MST++:

Hyperspectral imaging systems can be costly and complicated to use on a large scale in the agriculture industry. For this reason, the suggested framework utilizes MST++ (Multi-stage Spectral Transformer) for Spectral Reconstruction. The first step of the procedure involved converting regular RGB images of sugarcane leaves into realistic 31-band Hyperspectral Image Cubes within a visible wavelength spectrum. By using MST++, the algorithm will learn spectral correlations between RGB channels and Hyperspectral Signatures (HSS) and produce high dimensional spectral data. This method greatly diminishes dependence on hardware and allows for hyperspectral data to be generated that could be utilized for disease detection applications.

Wavelength Encoding:

After spectral reconstruction, the hyperspectral image cubes generated are subjected to pre-processing operations which include normalization, center patch extraction and tensor preparation. Centre patches that are small such as 5x5x31 and 7x7x31 are extracted to preserve local spectral-spatial information around the disease area. In order to improve the understanding of the spectral component of the data, a physics informed sinusoidal wavelength encoding mechanism has been incorporated into the framework. Because transformers have no natural understanding of wavelength progression, a custom wavelength encoding function maps spectral frequencies from approximately the visible light range of 400nm to 700nm. The wavelength aware embeddings obtained from this function are injected into the hyperspectral tensor streams, allowing the model to learn meaningful spectral ordering and frequency relationships.

Dual Model Evaluation Framework:

The methodologies in evaluating advanced deep learning models for hyperspectral disease detection include two models. The first model is a Spectral-Spatial Transformer (SST). This model takes spectral patches as sequential representations, employs multi-head self-attention techniques, and can capture long-range spectral dependencies, as well as characteristics present across multiple hyperspectral channels. Therefore, through attention-based learning, the model learns how to discriminate disease-related features from spectral channels.

The second model is Hybrid 3D-CNN + Transformer-based. This model includes a 3D convolutional layer that simultaneously extracts both spectral and spatial features from each hyperspectral cube. The feature maps are then flattened and passed in a lightweight transformer encoder for global context learning. Combining the ability of CNNs to perform local feature extraction with the ability of transformers to model long-range dependencies allows the proposed hybrid approach to improve disease classification robustness, as well as the feature representation.

Classification and Optimization:

In this phase, binary classification is performed to distinguish healthy sugarcane plants from those infected with Red Rot, the model passes feature representations (obtained from the second phase) through fully-connected classification layers, followed by softmax function of the final layer to generate a prediction. Model optimization techniques will include using the Adam optimizer and Cross-Entropy Loss function during training to ensure fast convergence. Early-stopping will monitor validation loss to prevent potential overfitting, and the model best-performing weights will be stored automatically. The proposed framework's performance will be evaluated using a variety of performance measurements, such as accuracy, precision, recall, F1 score, and confusion matrix analysis.

Mathematical Formulations:

The proposed framework incorporates mathematical operations for spectral reconstruction, wavelength-aware positional encoding, transformer attention learning, and optimization. Let the RGB image input be represented as $I_{RGB} \mathcal{E} \mathbb{R}^{H\times W\times 3}$, where H and W represent image height and width. The MST++ spectral reconstruction module transforms the RGB image into a hyperspectral cube $H_{HSI} \mathcal{E} \mathbb{R}^{H\times W\times 31}$ using spectral mapping functions:

$$H_{HSI} = f_{MST++}(I_{RGB}))$$

Where, f_{MST++} represents the learned spectral reconstruction function generating 31 wavelength bands. After reconstruction, normalization is applied as:

$$X_{norm} = \frac{X}{max(|X|)}$$

Where, X is the hyperspectral cube and Xnorm is the normalized spectral representation. To encode spectral sequence information, sinusoidal wavelength positional encoding is utilized:

$$PE(pos, 2i) = sin(\frac{pos}{10000^{\frac{2i}{d}}})$$

$$PE(pos, 2i+1) = cos(\frac{pos}{10000^{\frac{2i}{d}}})$$

Where, pos represents wavelength position, i is the spectral dimension index, and d denotes embedding dimension size.

The transformer model utilizes self-attention mechanisms to learn spectral–spatial dependencies from hyperspectral patches. Self-attention is computed as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\,(\frac{QK^T}{\sqrt{d_k}})\text{v}$$

where Q, K, and V represent Query, Key, and Value matrices respectively, and d_k denotes key dimension scaling. For classification, the extracted feature vectors are passed through fully connected layers followed by the SoftMax activation function:

$$p(y_i) = \frac{e^{z_i}}{\Sigma_j^c = 1\, e^{z_j}}$$

Where P(y_i) represents class probability, z_i denotes output logits, and C is the number of classes. Model optimization is performed using the Cross-Entropy Loss function:

$$L = -\sum_{i=1}^{c} y_i\, log(P(y_i))$$

Where y_i represents ground truth labels and P(y_i) denotes predicted probabilities. The Adam optimizer is employed to update model weights during training for improved convergence and stable learning. Experimental setup is provided in Table1.

Experimental Setup:

Table 1. Experimental Setup and Model Description

Category	Description
Programming Language	Python 3.13
Development Environment	Jupyter Notebook
Libraries & Tools	torch (PyTorch), torchvision, einops, numpy, scipy, h5py, scikit-learn (sklearn), matplotlib
Operating System	Windows based system with GPU
Hardware	AMD Ryzen 5 8645HS, NVIDIA RTX 4050, RAM = 32, Storage = 1TB
Validation Technique	Loss Monitoring, Early Stopping Technique, Accuracy Evaluation, Confusion Matrix Analysis, Precision Score, Recall Score, F1-Score Evaluation,
Important Hyperparameters	Learning Rate = 0.1, Batch Size, Number of Epochs = 50, Patch Size (5×5 and 7×7), Number of Spectral Bands (31 Bands), Activation Function (Softmax), Number of attention heads = 4, Dropout Rate = 0.1, Transformer Encoder Depth = 2 layers, Average Training Time 25–35 Minutes, Optimizer Type adam

Dataset Description:

In this research the dataset contains hyperspectral images of sugarcane leaves affected by red rot. Each dataset has 400 total images, 200 of which are images of healthy sugarcane leaves, while the remaining 200 are of diseased leaves. Each of those hyperspectral images contains 31 wavelengths, which enables detailed analysis of spectral abnormalities associated with diseases in sugarcane. In order to create an unbiased data collection for developing a model and assessing its performance, this hyperspectral dataset was created with a similar number of samples of both healthy and diseased sugarcane leaves (Table2). The RGB images of Red Rot-infected leaves from multiple sugarcane plants are collected. Then the RGB images were reconstructed into 31-band hyperspectral imagery using the MST++ Spectral Reconstruction Framework. These reconstructed hyperspectral images span the visible spectrum (400 nm - 700 nm) and will be kept in .mat file format for further processing.

Table 2. Dataset Description

Parameter	**Description**
Disease Type	Sugarcane Red Rot Disease
Data Type	Hyperspectral Image
File Format	.mat Files
Total Images	400 Images
Healthy Samples	200 Images
Affected Samples	200 Images
Spectral Bands	31 Bands
Wavelength Range	400 nm – 700 nm

Results and Discussion:

According to experimental findings, the transformer-based hyperspectral framework can be employed to classify Red Rot Disease on sugarcane leaves. Furthermore, the 31-band hyperspectral images utilized in this study indicate that the model could identify spectral spatial aspects from the data. The proposed system was established to be reliable for the early detection of disease via the evaluation of multiple performance measures, including accuracy, precision, recall, and F1-score.

Spectral–Spatial Transformer:

The classification performance is presented in Table 3

Table 3. Classification Performance of Proposed Framework

	Accuracy	Precision	Recall	F1-score
Healthy	0.65	0.65	0.67	0.66
Affected	0.65	0.66	0.63	0.64

According to the findings illustrated in Table 3, the Spectral-Spatial Transformer (SST) model performed with an overall accuracy of 65% when classifying red rot disease using sugarcane hyperspectral imagery. The SST model demonstrated a balanced result (precision) for healthy (0.65) and diseased (0.66) cane; however, the recall value for healthy (0.67) was slightly greater than the recall value for diseased (0.63). The F1-score values of 0.66 for healthy samples and 0.64 for diseased samples indicate moderate classification performance, and that there is potential for stable learning in the way of modelling spectral and spatial features. Using this method, the SST was able to exhibit good classification capability.

SST Confusion Matrix Analysis:

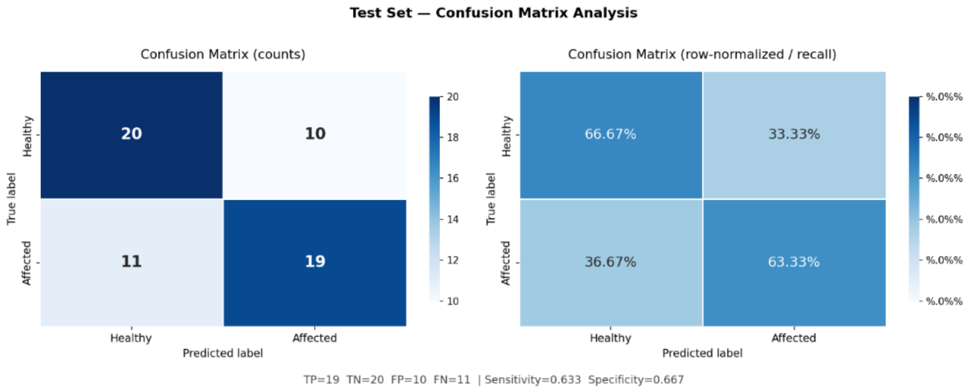

Figure 2. Confusion Matrix

The analysis of the SST model's performance in classification for both healthy and Red Rot affected sugarcane is outlined in the confusion matrix contained in figure 2. The model accurately classified 20 out of the 30 total healthy samples while incorrectly classifying 10 of them as having Red Rot. This results in a specificity of 0.667 for the healthy category of the model. Similarly, the model accurately classified 19 out of the 30 total Red Rot affected samples, while it incorrectly classified 11 samples as healthy. Thus, the sensitivity of the model for the Red Rot affected samples was 0.633. Based on the confusion matrix data, the classification capability of this model is moderate. The model performed slightly better on healthy samples than diseased samples.

Training and Validation Performance of SST Model:

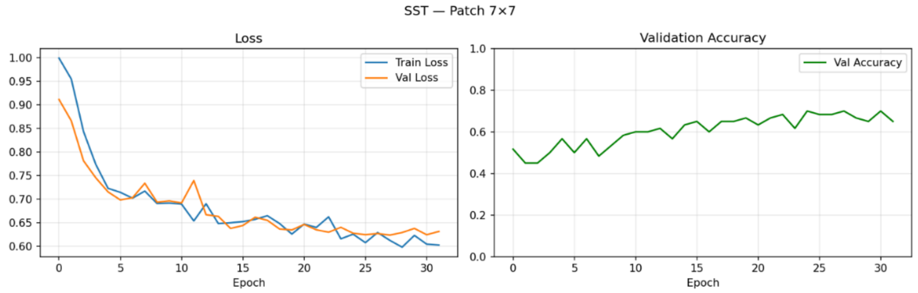

Figure 3. Training and Validation Performance

The Figure 3 shows the training and validation performance of the Spectral-Spatial Transformer (SST) model using a 7x7 patch size. The loss curve shows a decline in both the training and validation losses over the epochs, meaning that the model learned steadily without experiencing excessive model error during training. Likewise, the validation accuracy steadily increased from approximately 52% to approximately 70%, which provides evidence that the model learned progressively more meaningful spectral-spatial features for the classification of red rot disease without significant overfitting.

Hybrid 3D-CNN + Transformer:

The classification performance is presented in Table 4

Table 4. Classification Performance of Proposed Framework

	Accuracy	**Precision**	**Recall**	**F1-score**
Healthy	0.83	0.88	0.77	0.82
Affected	0.83	0.79	0.90	0.84

According to Table 4, the Hybrid 3D-CNN + Transformer model demonstrated good accuracy (83.33%) in classifying Red Rot disease in sugarcane hyperspectral images. Specifically, for identifying healthy sugarcane samples, the model had reasonable precision (0.88) and good recall (0.77), resulting in an F1-score of 0.82. The identification of healthy samples is reliable. The model performed well within the infected class, achieving precision (0.79), recall (0.90), and F1-score (0.84).

Hybrid Confusion Matrix Analysis:

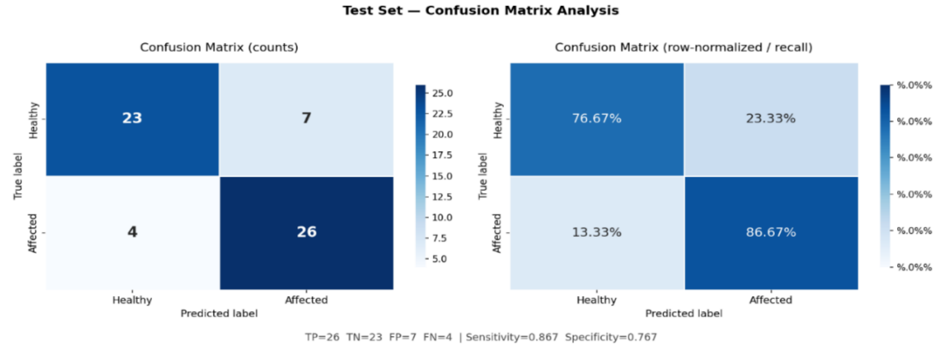

Figure 4. Hybrid Confusion Matrix

In Figure 4, you can see how well the Hybrid 3D-CNN + Transformer model is able to classify Red Rot disease in sugarcane using hyperspectral imagery data. For example, Of the 30 healthy samples, the model correctly identified 23 (but incorrectly identified 7 as being diseased) giving it a specificity rate of 76.67%. When dealing with the 30 diseased samples, the model correctly identified 26 of them (but incorrectly identified 4 as healthy). Therefore, it achieved a high sensitivity of 86.67%. The low number of false negatives and false positives indicates that the hybrid model captured spectral and spatial characteristics related to the disease.

Training and Validation Performance of Hybrid Model:

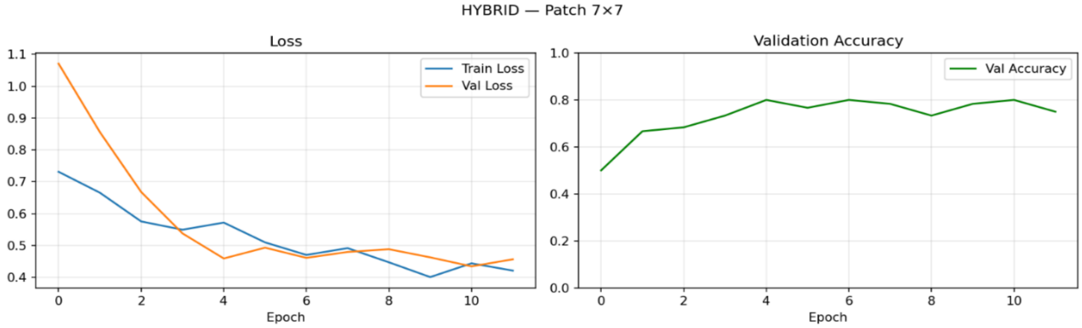

Figure 5. Training and Validation Performance of Hybrid Model

The performance of the Hybrid 3D + Transformer model can be found in the chart displayed in figure 5. It displays training performance and validation performance for a patch size of 7 x 7. As shown, training and validation loss decrease consistently throughout the epochs, illustrating an effective and improving rate of convergence in terms of the hybrid model's ability to learn. In terms of VAL Accuracy, it shows an upwards trend toward its final accuracy value over time and levels off at approximately 80%. These results demonstrate that the hybrid model effectively captures spectral-spatial characteristics within the hyperspectral domains, while remaining very generalizable, with very little to no overfitting.

Discussion:

Using a proposed hyperspectral disease detection framework, it shows how effectively transformer-based and hybrid deep learning models can identify Red Rot disease in sugarcane leaves. The SST (Spectral-Spatial Transformer) and Hybrid 3D-CNN + Transformer models successfully learned spectral-spatial features from the reconstructed 31-band hyperspectral image cubes of the dataset. The hyperspectral images also allow our models to capture extremely subtle spectral anomalies associated with disease infection, thereby supporting the early detection of disease prior to the appearance of severe visual symptoms.

The SST model produced moderate classification ability across the spectrum with 65% accuracy overall. This indicates that through the use of self-attention transformers, an ability to model long-range spectral dependencies, effectively is possible in hyperspectral remote sensing applications. The confusion matrix shows that the model showed moderate discrimination capability between healthy and affected samples; however, there were many false discriminations. The results indicate that while the ability of the SST architecture to learn spectral characteristics exists, its use as a single model to develop a significant amount of discrimination in very small patches of hyperspectral imagery may remain limited in its ability to provide a high discriminative quality. Due to overlapping spectral characteristics between healthy and diseased samples at an early stage of infection in sugarcane, some of the samples were classified incorrectly according to the confusion matrix analysis. There might be confusion in the spectral reconstruction created from the model due to inconsistencies which could lead to some false positives/negatives. These inconsistencies could be attributed to the small amount of data used to train the model. The Hybrid 3D-CNN + Transformer Model was able to reduce both the number of false positives and false negatives by capturing local spectral information (spectral-spatial) and global contextual information simultaneously.

The hybrid 3D-CNN + Transformer framework showed substantially better results than the SST model, with an overall accuracy of 83.33%, with higher precision, recall, and F1 scores. The hybrid architecture combines the local feature extraction capabilities of 3D convolutional layers with the global contextual learning capabilities of transformer self-attention mechanisms. The ability of the hybrid framework to achieve a high recall value for the affected class indicates that it can identify Red Rot infected samples accurately, which is critical for practical agricultural disease monitoring systems.

The training and validation performance curves confirmed the proposed hybrid framework is stable and effective for classification. The decreased training and validation loss over time indicates convergence and optimization have been achieved by the model after sufficient learning. The validation accuracy also converged to a stabilized accuracy of approximately 80%, suggesting good generalizability due to limited overfitting. The overall results support the integration of spectral–spatial (features extracted) with transformer attention (learning) to enhance the robustness of hyperspectral classification and improve disease recognition performance.

The proposed model demonstrates the effectiveness of integrating hyperspectral observations with deep learning and transformer attention mechanisms. The introduction of wavelength-aware positional encoding, and the reconstructions of hyperspectral image cubes through different spectral bands, all allowed the model to learn more representative datasets from reconstructed image cube datasets. The results demonstrated that hybrid architectures can offer reliable, interpretable and scalable methodologies for early identification of red rot disease on sugarcane plants, thus supporting the development of precision agriculture and supporting sustainable agricultural practices.

Implication of the Proposed Methodology:

Using hyperspectral imaging combined with a transformer-based deep learning approach, this methodology provides a novel and intelligent method for early detection of Red Rot Disease in sugarcane. The hybrid 3D-CNN + transformers architecture improves classification performance by capturing both local spatial features and non-local spectral dependencies from hyperspectral cubes through several processes, including spectral reconstruction, wavelength awareness in encoding methods and learning the combined spectral and spatial features at the same time to allow for early identification of disease-causing features prior to any significant symptom development.

The methodology will have major impacts on efficient farming and sustainable crop management. Early disease identification will allow farmers and agricultural specialists to decrease crop losses, increase sugarcane production, and improve their disease management plans. In addition to using explainable transformer attention mechanisms and hyperspectral imaging ability, both of which allow for greater interpretability of the model, will lead to adding more credibility to the entire system regarding its usefulness when making agricultural-related decisions in the "real world". This framework is a scalable and affordable choice for intelligent agricultural surveillance systems, therefore, it could be extended and could be beneficial for any other crop disease monitoring efforts.

Limitations:

This framework has some limitations, despite generating some interesting results. Firstly, this study used a limited amount of hyperspectral data, specifically, only 400 sugarcane leaves that were all collected from one location under controlled conditions, severely limiting how well the models can generalize outside of the data collection conditions. Secondly, because the data used in this study were generated from RGB images via MST++ as opposed to direct hyperspectral imaging via dedicated hyperspectral sensors, this may have led to some spectral reconstruction error, therefore causing very low spectral fidelity. Lastly, the computational resources needed for training and inferring the transformer-based models are substantial; thus, significant GPU acceleration is required to successfully train and infer these models. Importantly, at this point, the only classifications the framework performs are either 'healthy leaves' or 'leaves affected by red rot' and currently do not assess any other diseases that may affect sugarcane or the variability among environmental conditions (i.e., changing light, noise levels, and field-to-field variability). Because of these limitations, the researchers noted the need for larger datasets collected under realistic conditions, all hyperspectral acquisition methods being used in the framework, and improved multi-disease classification frameworks in future research.

Conclusion:

The researchers have developed a framework for analyzing hyperspectral images to detect Red Rot disease on sugarcane leaves early using a transformer model and hyperspectral image cubes made from 31 spectral bands to classify samples as healthy or affected. The framework includes: spectral reconstruction, preprocessing of data, a wavelength-aware positional encoding component for spectral data, and deep learning techniques to identify both spectral and spatial features of disease and to improve overall classification of disease. The first two tested networks were a Spectral-Spatial Transformer (SST) Model and a Hybrid 3D-CNN + Transformer Model to evaluate how network architectures influence disease classification performance.

The experiment demonstrated that the Hybrid 3D-CNN + Transformer combination exceeded the performance of a single SST model measured by accuracy, precision, recall and

F1 scores. The proposed hybrid framework is able to recognize both local spatial arrangements and also global spectral relationships, therefore it is able to provide better performance in detecting diseases as well as having higher generalizability. The overall results from the study show that hyperspectral imaging and transformer-based DL methods can be used to develop an intelligent method for investigating crop diseases, in addition this hybrid framework provides a promising tool for both precision agriculture and sustainable management of sugarcane diseases.

Future Directions:

There are multiple avenues of future research that can enhance the robustness and scalability of the proposed hyperspectral detection framework used to identify red rot disease through use of larger and more diverse sugar cane datasets that are collected from real-world environments. Adding additional spectral bands, environmental characteristics, and time-sequenced disease progression information can provide even greater accuracy and reliability over red rot disease's early detection. In addition, future researchers could investigate advanced transformer model architectures and low-complexity attention mechanisms to reduce the computational demands and enable real-time deployment for agricultural monitoring systems.

In addition, upcoming research efforts will include future studies that explore how Internet of Things (IoT), drones, and cloud-based Artificial Intelligence (AI) can create automated large-scale surveys for crops. Explainable Artificial Intelligence (XAI) could be modified to provide more accurate information about the spectral attention patterns and the disease biomarkers in a manner that improves the ability to interpret the XAI results. Additionally, the proposed framework could also be adapted to support the identification(s) of multiple diseases in sugarcane and other environmental stresses, thereby supporting the development of intelligent precision agriculture applications based on sustainable farming.

References:

[1] Dong Bao, Jun Zhou, “Early detection of sugarcane smut and mosaic diseases via hyperspectral imaging and spectral-spatial attention deep neural networks,” *J. Agric. Food Res.*, vol. 18, p. 101369, 2024, doi: https://doi.org/10.1016/j.jafr.2024.101369.

[2] Xiaohui Huang, Yunfei Zhou, “SS-TMNet: Spatial–Spectral Transformer Network with Multi-Scale Convolution for Hyperspectral Image Classification,” *Remote Sens.*, vol. 15, no. 5, p. 1206, 2023, doi: https://doi.org/10.3390/rs15051206.

[3] Z. Zhong, Y. Li, L. Ma, J. Li, and W. S. Zheng, “Spectral–Spatial Transformer Network for Hyperspectral Image Classification: A Factorized Architecture Search Framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2021.3115699.

[4] Malithi De Silva, Dane Brown, “Multispectral Plant Disease Detection with Vision Transformer–Convolutional Neural Network Hybrid Approaches,” *Sensors*, vol. 23, no. 20, p. 8531, 2023, doi: https://doi.org/10.3390/s23208531.

[5] Hanxiang Wang, Tri Hai Nguyen, “PD-TR: End-to-end plant diseases detection using a transformer,” *Comput. Electron. Agric.*, vol. 224, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0168169924005143

[6] Anuruk Prommakhot, Jakkrit Onshaunjit, Wichian Ooppakaew, Grianggai Samseemoung, Jakkree Srinonchat, “Hybrid CNN and Transformer-Based Sequential Learning Techniques for Plant Disease Classification,” *IEEE Access*, 2025, [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11072169

[7] Manh Tuan Do, Manh Hung Ha, “Toward improving precision and complexity of transformer-based cost-sensitive learning models for plant disease detection,” *Front. Comput. Sci.*, vol. 6, 2024, doi: https://doi.org/10.3389/fcomp.2024.1480481.

[8] Ali Zia, Jun Zhou, Muyiwa Olayemi, “Determining Mosaic Resilience in Sugarcane Plants using Hyperspectral Images,” *arXiv:2501.16700*, 2025, [Online]. Available: https://arxiv.org/abs/2501.16700

[9] Aryan Kumar, Gaurav Saini, “A Comparative Study of Deep Learning Approaches for

Early Detection of Sugarcane Diseases," *Procedia Comput. Sci.*, vol. 260, pp. 182–190, 2025, doi: https://doi.org/10.1016/j.procs.2025.03.192.

[10] Lan Zhang, Yang Wang, "A multi-range spectral-spatial transformer for hyperspectral image classification," *Infrared Phys. Technol.*, vol. 135, p. 104983, 2023, doi: https://doi.org/10.1016/j.infrared.2023.104983.

[11] Fernanda Rodrigues Silva, Mário Lúcio V.de Resende, "Molecular Characterization and Pathogenicity of Colletotrichum falcatum Causing Red Rot on Sugarcane in Southern Florida," *J. Fungi*, vol. 10, no. 11, p. 742, 2024, doi: https://doi.org/10.3390/jof10110742.

[12] R. Viswanathan, R. Selvakumar, K. Manivannan, R. Nithyanantham, and K. Kaverinathan, "Behaviour of Soil Borne Inoculum of Colletotrichum falcatum in Causing Red Rot in Sugarcane Varieties with Varying Disease Resistance," *Sugar Tech*, vol. 22, no. 3, pp. 485–497, Jun. 2020, doi: 10.1007/S12355-020-00800-7.

[13] R. Viswanathan and R. Selvakumar, "Varietal Breakdown to Red Rot in Sugarcane Revealed by Comparing Two Colletotrichum falcatum Inoculation Methods," *Sugar Tech 2020 226*, vol. 22, no. 6, pp. 1063–1075, Jul. 2020, doi: 10.1007/S12355-020-00855-6.

[14] A. Chandra, A. T. Keizerweerd, Y. Que, and M. P. Grisham, "Loop-mediated isothermal amplification (LAMP) based detection of Colletotrichum falcatum causing red rot in sugarcane," *Mol. Biol. Rep.*, vol. 42, no. 8, pp. 1309–1316, Aug. 2015, doi: 10.1007/S11033-015-3875-9.

[15] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis", [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0168169917311742?via%3Dihub

[16] Edna Chebet Too, Li Yujian, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agric.*, vol. 161, pp. 272–279, 2019, doi: https://doi.org/10.1016/j.compag.2018.03.032.

[17] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artif. Intell. Rev. 2020 538*, vol. 53, no. 8, pp. 5455–5516, Apr. 2020, doi: 10.1007/S10462-020-09825-6.

[18] D. Wang *et al.*, "Early Detection of Tomato Spotted Wilt Virus by Hyperspectral Imaging and Outlier Removal Auxiliary Classifier Generative Adversarial Nets (OR-AC-GAN)," *Sci. Reports 2019 91*, vol. 9, no. 1, pp. 1–14, Mar. 2019, doi: 10.1038/s41598-019-40066-y.

[19] Muhammad Shafay, Taimur Hassan, Muhammad Owais, Irfan Hussain, Sajid Gul Khawaja, Lakmal Seneviratne & Naoufel Werghi, "Recent advances in plant disease detection: challenges and opportunities," *Plants Methods*, vol. 21, no. 140, 2025, [Online]. Available: https://link.springer.com/article/10.1186/s13007-025-01450-0

[20] I. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, p. 30, 2017.

[21] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-Excitation Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020, doi: 10.1109/TPAMI.2019.2913372.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-December, pp. 770–778, Dec. 2016, doi: 10.1109/CVPR.2016.90.

[23] S Aswani, Ashish Sinha, "Early diagnosis of Sugarcane leaf diseases through CNN and Vision Transformer hybrid model," *Researchgate*, 2025, doi: 10.21203/rs.3.rs-7232200/v1.