



SugarcaneGAN: A U-RSwinT and StyleGAN3-Driven Deep Learning Framework for High-Accuracy Sugarcane Leaf Disease Detection and Classification

Ramesh S. Lavhe,^{1,*} Satyajit Pangaonkar,¹ Parikshit N. Mahalle,² Sarika Vasantrao Bodake,³ Tushar Jadhav,² Deepali S. Jadhav,² Sumit Arun Hirve,⁵ Snehlata Wankhade,⁴ Dattatray G Takale² and Arvind Jagtap⁶

Abstract

Early and accurate diagnosis of sugarcane leaf diseases is essential for sustainable agriculture, improved crop productivity, and minimizing industry losses. Manual inspection methods are often inefficient, labor-intensive, and prone to inconsistency, especially in large-scale farming contexts. SugarcaneGAN refers to our custom generative model designed to augment the dataset with high-quality synthetic diseased leaf images, addressing class imbalance and improving model generalization. This study introduces SugarcaneGAN, a comprehensive deep learning framework that integrates advanced image preprocessing, segmentation, generative modeling, and classification techniques to deliver high-accuracy sugarcane leaf disease detection. The pipeline begins with the acquisition of RGB images, followed by Contrast Limited Adaptive Histogram Equalization (CLAHE) to enhance contrast, median filtering for noise suppression, and a transformation to HSV color space to emphasize disease-specific chromatic patterns. For precise segmentation of sugarcane leaves in cluttered environments, the framework utilizes U-RSwinT a U-Net architecture enhanced with a Residual Swin Transformer backbone leveraging hierarchical design and self-attention mechanisms to capture both global context and local features. The Segment Anything Model (SAM) is optionally incorporated for zero-shot segmentation, thereby improving flexibility and robustness. To address dataset limitations such as class imbalance and low variability, a custom generative module named SugarcaneGAN is developed using StyleGAN3 and DiffusionGAN to synthesize high-fidelity diseased leaf images. These augmented samples enhance model generalization during training. Feature extraction is conducted via the ConvNeXt architecture, further enriched by Gray Level Co-occurrence Matrix (GLCM) texture descriptors that capture lesion structure. The classification stage utilizes vision transformer-based models, including MaxViT, DeiT, and ViT-B16, to identify disease type, assess severity, and localize affected regions. This integrated approach enables a scalable, real-time solution for intelligent sugarcane disease monitoring, aligning with the broader goals of precision agriculture, smart farming, and global food security. Experimental evaluation demonstrates SugarcaneGAN achieving over 98% accuracy and F1-score across six disease classes. The model also records a highest AUC of 0.97 and average precision (AP) of 0.90, confirming its superior generalization and reliability under class imbalance.

Keywords: Sugarcane leaf disease; U-RSwinT; StyleGAN3; Deep learning; Image classification.

Received: 26 September 2025; Revised: 15 November 2025; Accepted: 28 November 2025

Article type: Research article.

1. Introduction

Agriculture remains one of the pillars of the economy, particularly for crops such as sugarcane, which are extensively grown in the tropics and subtropics. Sugarcane is a vital industrial crop used to produce sugar, ethanol, paper, and biofuel; therefore, the health and production of sugarcane are issues of national concern.^[1] Nevertheless, sugarcane crops are highly susceptible to various foliar diseases caused by

different fungal, bacterial, and viral pathogens. When these diseases are not detected and addressed promptly, they can lead to significant production losses and poor crop quality. Leaf diseases are evident in their symptoms, which are often overlooked or misunderstood due to the minor variations in symptoms and disease cases. The manual method of inspection is the most commonly used approach, but it is inefficient, subjective, and time-consuming, especially in large-scale monitoring.^[2] The increasing interest in precision agriculture has led to an investigation into the potential of new

¹School of Engineering, Avantika University, Ujjain, Madhya Pradesh, 456006, India

technologies, such as computer vision and artificial intelligence, as tools for disease detection and management. All these technologies provide throughput, non-invasive, and scale-up data to magnify field diagnostics. As more people gain access to relatively inexpensive mobile devices, drones, and edge computing, there is an increase in the possibility of incorporating disease recognition systems based on AI into typical agricultural practices.^[3] Building an innovative system to identify diseases in sugarcane leaves is also a challenging task because there are many varieties of these diseases, varying environmental conditions, and image variability that make this approach difficult. Therefore, automated systems that are computationally efficient yet possess high classification accuracy wherever possible are sorely needed to facilitate sustainable farming.^[4]

In some parts of the world, particularly in tropical and subtropical regions, sugarcane crops are highly cultivated, and agriculture remains a central component in supporting many economies today. Sugar cane is an industrial crop that is preferred in the production of sugar, ethanol, paper, and biofuels; this makes the health and yield of sugar cane of national concern. Nonetheless, sugarcane plants are highly susceptible to most foliar diseases caused by fungal, viral, and bacterial agents.^[5] Unless these diseases are identified and controlled promptly, they may cause plant species to fail to produce high yields, resulting in poor-quality crops. Leaf diseases, especially, can exhibit visual symptoms that are often overlooked or misdiagnosed due to slight differences in the types of diseases. The most widely used approach, manual inspection, is time-consuming, subjective, and costly, as it requires extensive monitoring. The emerging demand for precision agriculture has triggered the investigation of high-tech solutions, such as computer vision and artificial intelligence, to identify diseases. The technologies provide high-throughput, non-invasive, and scalable methods to enhance field diagnostics. As more people gain access to cost-effective mobile devices, drones, and edge computing, it is even possible to integrate AI-driven disease recognition systems into daily agricultural practices. Nonetheless, the ability to construct an intelligent system that can consistently identify sugarcane leaf diseases is a challenge due to the variation in disease symptoms, environmental conditions, and

image quality.^[6] Therefore, there is an urgent need to implement automated systems that combine computational efficiency with a high level of classification, thereby favoring sustainable farming methods.

Traditional and modern methods have been used in plant leaf disease detection, including statistical learning, image processing, and machine learning models. The earlier techniques, such as SVM, k-NN, and decision trees, relied widely on features that were manually designed, or handcrafted, including colour histograms, texture, and shape descriptors. Although these approaches work well in well-controlled settings, they are not scalable, noise-averse, and environment-responsive. With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), there has been a significant improvement in disease classification accuracy, yielding hierarchical features that facilitate advanced classification. However, through automatic learning, hierarchical features of the raw images were obtained. Nonetheless, CNNs tend to be highly data-demanding to achieve consistent performance, and they rarely perform well in agricultural data.

Additionally, they are susceptible to difficulties in cases involving visually similar diseases or varying illuminations, angles, and backgrounds that images in the field often exhibit. Despite synthetic data augmentation methods being used to expand the data size, conventional augmentation mechanisms, such as flipping, rotating, or creating colour jittering, are ineffective in capturing the intricate characteristics of features. Transformer-based models have shown promise in other fields but tend to be inefficient or overfit in small agricultural training sets. Therefore, despite these developments, traditional methods still do not provide robust, generalized, and scalable disease type detection systems that can be applied to precision agriculture.

The proposed SugarcaneGAN framework presents a novel end-to-end solution for automated sugarcane leaf disease detection, combining contrast-enhanced preprocessing, advanced segmentation with transformer-based U-RSwinT, and zero-shot capability using SAM. Synthetic augmentation using StyleGAN3 and DiffusionGAN addresses data imbalance and enhances training diversity. Feature extraction through ConvNeXt and GLCM ensures a rich representation of color-texture patterns, which are critical for accurate diagnosis. Classification via ViT-based models delivers robust multi-class identification with high interpretability. This cohesive pipeline significantly outperforms conventional models, making it ideal for deployment in real-time precision agriculture systems. The primary benefit of the study is as follows,

- Integrated Deep Learning Pipeline for Disease Detection: This study introduces SugarcaneGAN, a unified and end-to-end framework combining advanced preprocessing (CLAHE, median filtering, HSV transformation), transformer-based segmentation (U-RSwinT), and robust classification (ViT-family models) to ensure high-accuracy sugarcane leaf disease

²Vishwakarma Institute of Technology, Pune, Maharashtra, 411037, India

³Padamabhooshan Vasantdada Patil Institute of technology, Bavdhan, Pune, 411021, India

⁴Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, 440008, India

⁵Department of Computer Science and Engineering, MIT Art Design and Technology University, Rajbaug Loni Kalbhori, Pune, Maharashtra, 412201, India

⁶Computer Engineering Department, Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering & Technology, Baramati, Maharashtra, 413133, India

*Email: ramesh.lavhe@gmail.com (R. S. Lavhe)

detection in complex agricultural environments.

- **Novel Generative Module for Data Augmentation:** To address data scarcity and class imbalance, a customized generative framework blending StyleGAN3 and DiffusionGAN is developed, producing diverse and realistic synthetic images of diseased leaves. These augmentations significantly enhance model robustness and generalization.
- **Multi-level Feature Extraction and Classification Strategy:** By integrating ConvNeXt for deep feature learning and GLCM for lesion texture analysis, followed by Transformer-based classifiers (MaxViT, DeiT, ViT-B16), the system accurately classifies disease types and severity while enabling lesion localization, supporting intelligent and real-time agricultural decision-making.

The detection of plant diseases has become increasingly important in the modern world due to its significant impact on agricultural yields. The conventional CNN-based models have gained popularity; however, when utilized, they struggle to handle complex disease patterns or deal with small datasets. More modern developments utilize transformer structures and GANs to achieve better feature extraction and data augmentation purposes. This section discusses and evaluates

some related work, highlighting existing gaps and justifying the suggested SugarcaneGAN framework.

It has been demonstrated by Miryala and Rasane^[7] that Vision Transformer-based models outperform CNNs in sugarcane leaf disease classification, as they are more adept at identifying long-term interdependencies and patterns. They explore the increasing importance of transformer-based models in plant pathology. Rajput *et al.*^[8] modeled MLTSDC, a multi-level attention transformer-based disease detection classifier for sugarcane disease, which utilizes deep attention levels in its disease classification. They confirm the potential of hierarchical attention mechanisms in perceiving subtle differences between disease classes.

Another non-visual approach is by J-L. Ong *et al.*^[9] which involves the combination of visible and near-infrared spectroscopy variables with the help of the continuous wavelet transform-based spectrograms, adding in deep learning. In this approach, detection accuracy was reported to be high, indicating the potential of multimodal sensing in crop disease detection. Biswas *et al.*^[10] proposed ResNet, an ensemble deep learning network that combines several backbone CNNs to maximize sugarcane leaf disease classification. The study

Table 1: Problem formulation of the conventional techniques.

Citation	Methodology Used	Result of Papers	Research Gap
Miryala and Rasane ^[7]	Vision Transformer-based classification	Achieved 94.6% classification accuracy on the sugarcane leaf dataset with better generalization	Did not integrate synthetic data generation or hybrid attention-augmentation frameworks
Rajput <i>et al.</i> ^[8]	MLTSDC – Multi-Level Attention Transformer	Reported 95.3% accuracy using hierarchical attention features for fine-grained classification	No integration of generative or self-supervised pre-training components
J-L. Ong <i>et al.</i> ^[9]	NIR + Visible spectroscopy with wavelet transform and deep learning	Obtained 93.1% accuracy with multimodal spectrogram learning	Non-visual sensing is difficult to deploy in large-scale visual-only field systems
Biswas <i>et al.</i> ^[10]	ResNet-based ensemble model combining multiple CNNs	Improved classification robustness to 92.8% accuracy	Did not adopt attention-based models or GAN-based augmentation techniques
Rabbany <i>et al.</i> ^[11]	DVMNet – Ensemble of CNN classifiers	Recorded 91.4% accuracy; lacked generalization under varying lighting or noise conditions	Limited dataset size and no synthetic image generation
Chaudhary <i>et al.</i> ^[12]	Random Forest, SVM, and Decision Trees on handcrafted features	Achieved 84.7% accuracy with interpretable results	Less effective for high intra-class variability; lacked automated feature learning
Kong <i>et al.</i> ^[13]	Biological/epidemiological review of leaf scald disease	Provided qualitative insights into sugarcane disease lifecycle and ecological spread	No computational model; limited impact on real-time disease detection frameworks
Hambir and Sandhya ^[14]	Drone-based imagery + AI for brown-spot detection	Achieved 89.2% accuracy from aerial images in field trials	No integration of fine-grained close-up leaf analysis or mobile ground truth confirmation
Azghadi <i>et al.</i> ^[15]	Robotic weed detection and intervention systems	Demonstrated 88.3% detection accuracy in real-time with robotic platforms	Focused on weed species; irrelevant to sugarcane-specific foliar disease diagnosis
Askale <i>et al.</i> ^[16]	Light-weight CNN deployed on mobile for maize leaf disease detection	Achieved 90.5% accuracy with real-time inference on mobile edge devices	Did not explore transformer-based mobile models or sugarcane domain adaptation

highlights the success of a model ensemble in improving the overall reliability of the prediction. Table 1 show Problem found in exiting system.

DVMNet is an ensemble of CNN classifiers developed by Rabbany *et al.*^[11] to detect sugarcane leaf diseases. Their structure was very accurate, but had not yet undergone generalization due to the limited quantity of data and the failure to augment it. Chaudhary *et al.*^[12] employed machine learning algorithms to conduct a predictive analytics exercise aimed at identifying sugarcane diseases. Although their solution was well-explained and easy to replicate in reality, it was not as precise and sophisticated as the deep learning solution. The comprehensive review of sugarcane leaf scald disease given by Kong *et al.*^[13] focuses on discussing the pathogenic process and eco-effective management strategies of that disease. Their study, though not a topic of AI interest, still provides important biological contextualization in detecting the target disease and training the models.

Hambir and Sandhya^[14] focused on investigating the application of AI and imagery captured by drones to detect the brown-spot leaf illness on sugarcane. Their article highlights the roles of aerial data buying and remote sensing in intelligent farming systems. The authors noted that accurately identifying the location of diseases is crucial when employing robotic spot-spraying in sugarcane fields to minimize herbicide use (Azghadi *et al.*^[15]). The paper describes the importance of accurate detection systems for automated interventions, with a primary focus on weed control. Askale *et al.*^[16] introduced a deep CNN model for mobile devices to classify maize leaf diseases, indicating that the model is scalable and can be used in real-time. Their solution promotes the concept of using light AI models on mobile phones or edge devices for field diagnosis.

The structure of this research is as follows: Section 2 entails the literature review, an overview of current methods in plant disease detection, and the present shortcomings in segmentation, classification, and augmentation strategies. Section 3 presents the proposed SugarcaneGAN framework, describing how to combine the U-RSwinT model as a semantic segmentation approach and StyleGAN3 as the model for generating synthetic data, as well as the classification pipeline that completes the end-to-end process. In Section 4, the experimental setup and performance analysis of SugarcaneGAN, compared to conventional CNNs, transformer-based models, and other GAN-enhanced models, are discussed in terms of accuracy, precision, recall, and F1-score. Lastly, the study is summarised in Section 5, and a future direction for the research is outlined.

2. Computational framework for disease classification

The proposed SugarcaneGAN framework follows a systematic and modular pipeline, designed to ensure robust detection of sugarcane leaf diseases. The process begins with the acquisition of high-resolution RGB images of sugarcane leaves under natural field conditions. These images undergo

preprocessing using CLAHE to enhance local contrast and median filtering to suppress noise while preserving edge details. Subsequently, the RGB images are transformed into the HSV color space to accentuate chromatic disease symptoms, facilitating better distinction between healthy and infected regions. For leaf segmentation, the model utilizes U-RSwinT, a U-Net variant incorporating Residual Swin Transformer blocks. This architecture leverages hierarchical attention and residual connections to capture both global context and fine-grained local features, enabling accurate segmentation in cluttered and variable backgrounds. An optional module, the SAM, is included for zero-shot segmentation scenarios, enhancing the model's adaptability and generalization across different datasets. The proposed architecture is presented in Fig. 1.

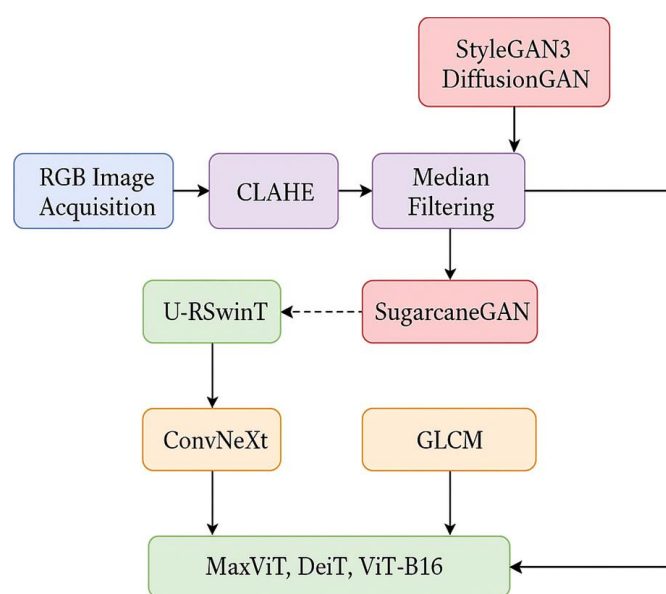


Fig. 1: Proposed system architecture.

To address the common challenge of class imbalance and limited variability in agricultural datasets, a custom generative module is introduced. This module combines StyleGAN3 and DiffusionGAN to synthesize high-fidelity images of sugarcane leaves exhibiting various disease symptoms. These synthetic images are then used to augment the training dataset, improving the diversity and balance of the input samples. After segmentation and augmentation, feature extraction is conducted using ConvNeXt, a modern convolutional backbone that excels at learning hierarchical representations from visual data. This deep feature set is enriched with GLCM descriptors to incorporate textural information associated with disease lesions, such as roughness, homogeneity, and entropy. Finally, the classification stage utilizes advanced vision transformer models—specifically, MaxViT, DeiT, and ViT-B16—to predict disease types, assess severity, and localize affected regions. The output includes class labels, severity scores, and annotated leaf images for real-time monitoring and analysis. This holistic approach not only ensures high classification accuracy but also contributes to sustainable crop

management through precision agriculture and intelligent disease monitoring.

2.1 Image preprocessing using CLAHE, median filtering, and HSV transformation

Image preprocessing is a crucial step in the proposed SugarcaneGAN framework, designed to enhance the quality and discriminative features of sugarcane leaf images prior to segmentation and classification. The preprocessing pipeline includes three main stages: CLAHE, Median Filtering, and HSV Color Space Transformation.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

CLAHE enhances local contrast in images by processing small contextual regions (tiles) rather than the entire image. It prevents over-amplification of noise by applying a clip limit to the histogram values. For each tile, the histogram is calculated and clipped if any bin exceeds the predefined threshold.^[17] In Eq. (1), The cumulative distribution function (CDF) is then used to map the input pixel intensity I_x, y to an output pixel intensity I_{CLAHE} using,

$$I_{CLAHE}(x,y)=CDF_{clipped}(I(x,y)) \times (L-1) \quad (1)$$

here, $CDF_{clipped}$ refers to the clipped cumulative histogram, and L represents the number of gray levels. This enhances contrast in localized areas, making subtle disease patterns more distinguishable.

Median Filtering

Median filtering is applied to reduce salt-and-pepper noise and preserve edges, which is especially important for maintaining lesion boundaries. In Eq. (2), The median filter processes each pixel by replacing it with the median of neighboring pixel intensities within a window W (typically 3×3 or 5×5):

$$I_{median}(x,y)=\text{median}\{I(i,j)|(i,j) \in W(x,y)\} \quad (2)$$

This non-linear filtering technique is effective in removing noise without significantly blurring lesion details.

HSV Color Space Transformation

After denoising and contrast enhancement, the image is transformed from RGB to HSV (Hue, Saturation, Value) color space to highlight disease-related chromatic patterns^[18] better. The HSV representation decouples intensity (Value) from color information (Hue and Saturation), enabling better differentiation of symptoms such as yellowing, browning, or necrosis. In Eq. (3, 4), Show the conversion from RGB to HSV is computed via:

$$H = \begin{cases} 0, & \Delta=0 \\ 60^\circ \times \left(\frac{G-B}{\Delta} \bmod 6 \right) & C_{\max}=R \\ 60^\circ \times \left(\frac{B-R}{\Delta} + 2 \right) & C_{\max}=G \\ 60^\circ \times \left(\frac{R-G}{\Delta} + 4 \right) & C_{\max}=B \end{cases} \quad (3)$$

$$S = \begin{cases} 0 & C_{\max} \\ \frac{\Delta}{C_{\max}} & \text{otherwise} \end{cases}; \quad V = C_{\max} \quad (4)$$

here, $C_{\max} = \max(R, G, B)$, $C_{\min} = \min(R, G, B)$, $\Delta = C_{\max} - C_{\min}$. This transformation highlights subtle color variations in diseased areas, enabling improved performance in downstream segmentation tasks.

2.2 Segmentation via U-RSwinT and segment anything model (SAM)

Accurate segmentation of sugarcane leaves is essential for isolating diseased regions from complex backgrounds. This study employs a hybrid segmentation strategy combining the strengths of U-RSwinT and the SAM. U-RSwinT integrates the powerful hierarchical attention of the Swin Transformer with the U-Net's encoder-decoder structure to capture both local and global features. Additionally, SAM is incorporated as an optional zero-shot segmentation tool, enhancing adaptability across unseen environments and datasets. The detailed architecture will be known as the U-RSwinT, which can be referred to as a hybrid model that incorporates a residual U-Net with attention-enhanced Skip connections and a 3D Swin Transformer in the bottleneck. The network design is based on a 3-layer learning Cube and a 4-position residual U-Net. Every encoder block first utilizes 3D max pooling to halve the spatial size and then employs two chains of group normalization, leaky ReLU activation, and 3D convolution. A subordinate linkage normalizes and convolves the input and then combines it with the output of the two convolutions, and in effect, resolves the problem of the vanishing gradient. Instead, the encoder doubles the number of channels per stage (16 to 32, 32 to 64, 64 to 128), allowing it to extract higher-order features. Central to the model is the 3D Swin Transformer block, which encodes the output of the 128-channel encoder using a single transformer layer with six attention heads, a patch size of $2 \times 2 \times 2$, and an embedding dimension of 96. This element works effectively in modelling the long-range dependencies through a shifted window self-attention. It is followed by a $3 \times 3 \times 3$ convolution that doubles the channel depth to 256, matching the input of the decoder.

In Fig. 2, The decoder employs 3D transposed convolutions to upsample feature maps, and each convolution block is preceded by attention blocks that further focus on relevant details by combining the respective outputs of the encoder. These mini-selected maps undergo residual double convolution, comprising group normalization layers, leaky ReLU, and convolutional layers, ultimately resulting in a $1 \times 1 \times 1$ convolution to produce the final 3D segmentation mask. Residual learning, attention mechanism, and transformer-based modeling of the U-RSwinT architecture are successfully integrated, resulting in a versatile and expandable structure with approximately 6.19 million parameters, which enables high-accuracy 3D medical image segmentation.

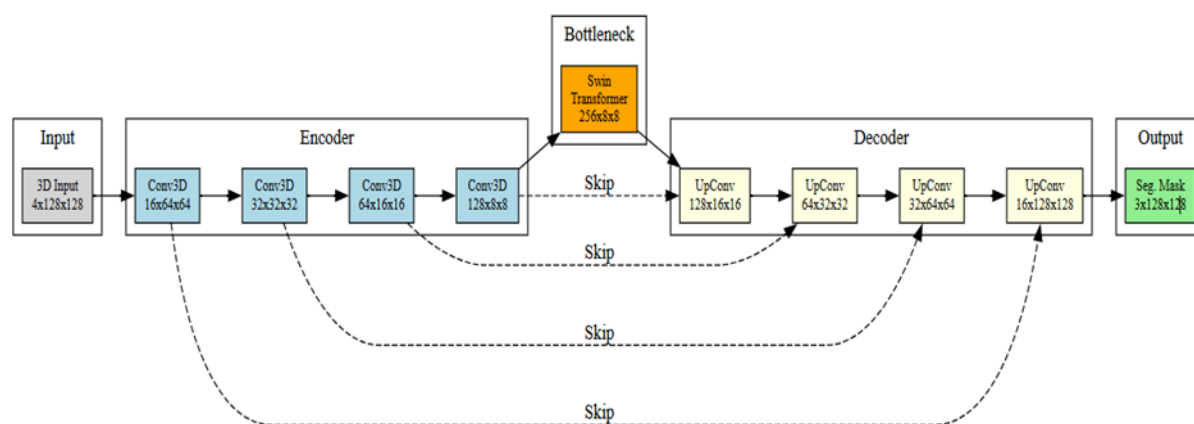


Fig. 2: Architecture of U-RSwinT.

2.2.1 Residual U-Net

U-Net is a CNN architecture that segments images, typically represented in a U-shape form. It comprises an encoder that derives hierarchical features from the high-resolution input images and a decoder that up-samples these features to produce segmentation masks. Between them, skip links are used to connect the layers, aiming for equivalence. The issue of vanishing gradients is a significant problem in deep networks, such as the U-Net. The gradients in earlier layers become so small that the change in weights after backpropagation becomes minuscule, and the amount of learning becomes relatively small. The problem can be addressed through a residual U-Net structure, which incorporates residual blocks — shortcuts between layers — that enable a better flow of gradients and facilitate the network's ability to learn the process. This design is further developed in the proposed SugarcaneGAN framework, where Swin Transformer blocks are incorporated into the residual U-Net to introduce U-RSwinT (where 'U' is used to denote both local feature extraction and the global context modeling module), which is efficacious and robust in segmenting sugarcane leaf disease. Eq. (5) provides the output in the block.

$$y = F(x, \{W_i\}) + x \quad (5)$$

In this case, y is the output of the residual block, and x is the input of the residual block. The residual block is a function of $F(x, \{W_i\})$, which is usually a feedback of convolutional layers, and W_i are the weights of the i th layer stage in a unit of the residual. With this structure, there is a dimension of gradients passing directly across the layers of the network through shortcut connections. Consequently, training will be quicker and regular when working with deep neural networks.

2.2.2 Swin transformer

Although CNNs have demonstrated enormous efficiency in solving diverse visual problems, they are significantly challenged in representing global contextual relationships due to the localized receptive fields of the CNN architecture. The weakness of the vision Transformers is circumvented by employing self-attention, which captures long dependencies in

the input data, hence utilizing long-range features to express them. However, Standard Vision Transformers are computationally intensive to apply to huge resolution images. To mitigate this challenge, the Swin Transformer architecture design has emerged based on the idea of a practical hierarchical simplification of Vision Transformers. The primary focus of such a model is on its scalability with image size and computational efficiency. The Swin Transformer begins by breaking the input picture into linearized patches, unifying the embedding of these patches, and then resizing and grouping them into blocks for a transformer. The blocks include normalization, attention mechanisms, and a multi-perceptron. Amazingly, the W-MSA is a scheme that confines computational complexity to local windows, which are non-overlapping. Under this scheme, the complexity is significantly reduced compared to the quantity of patches, transitioning from quadratic to linear complexity. This local-global attention mechanism enables Swin Transformer to be efficient in processing large images, retaining not only structural data but also local details, which makes it highly applicable to segmentation tasks such as detecting sugarcane leaf diseases, as seen in this example of illness detection. In Eq. (6, 7) show the MSA's quadratic term,

$$(MSA) = 4hwC^2 + 2(hw)^2c \quad (6)$$

$$\Omega(W-MSA) = 4hwC^2 + 2M^2hwC \quad (7)$$

As the number of patches hw rises, the original MSA's quadratic term, $(h)^{w^2}c$, becomes costly, particularly for high-resolution images. In the Swin Transformer architecture, the linear complexity term $2M^2hwC$ associated with Window Multi-Head Self-Attention (W-MSA) offers significant scalability advantages over standard self-attention. The term does not depend on the number of patches because the window size M is fixed; thus, the model is more computationally efficient for high-resolution inputs.^[19] To extend the global context modeling, the architecture involves the integration of a second subunit, Shifted Window Multi-Head Self-Attention (SW-MSA). This applies a cyclic shift of the input and then conducts self-attention across the shifted windows. It is a

method that allows cross-window interaction and enables the learning of more extensive spatial relationships within the image. Moreover, Swin Transformers introduce positional relative bias in attention to Transformers, allowing them to learn positional relationships between patches even more effectively. In Eq. (8), This two-attention approach, which combines local attention with effective global context extraction, enables the Swin Transformer to be specifically effective in dense prediction tasks, such as image segmentation and disease detection in agricultural photos.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \quad (8)$$

The Swin Transformer model uses Queries (Q), Keys (K), and then Values (V) based on input patches to calculate attention weights with a dimension d of the vectors. Hence, the bias of relative positions between patches is defined as a matrix B , which encodes the relative position relationship among the patches. Each transformer block features a multi-layer perceptron (MLP) component, comprising two fully connected layers with an intermediate non-linear activation function, enabling the computation of complex non-linear relationships between features. Layer normalization is introduced before and after multi-head self-attention (MSA) and MLP modules to stabilize training and achieve better convergence. Additionally, residual connections are introduced, which allow the original input to be directly added to the output, thereby enhancing the ability to avoid the vanishing gradient issue and facilitating the preservation of information throughout the entire network.^[20] Additionally, Swin Transformer employs a hierarchical patch merging approach, where neighboring patches are concatenated along the channel dimension and then downsampled at a rate of N . This top-down fusion not only alleviates computation but also enhances the model's capabilities in responding to contextual information at various levels.

2.3 Synthetic data generation with StyleGAN3 and ddiffusionGAN

To overcome the challenges of limited and imbalanced datasets in sugarcane disease detection, this study incorporates a dual generative approach using StyleGAN3 and DiffusionGAN. StyleGAN3 enables the generation of high-resolution, realistic sugarcane leaf images with diverse disease patterns through style-based modulation. Complementarily, DiffusionGAN utilizes iterative denoising to generate fine-grained synthetic samples that closely replicate real-world variations.^[21] Together, these models significantly enrich the training data, improving the robustness and generalization of downstream deep learning classifiers. The power of the proposed segmentation framework on the U-RSwinT is high due to the vast amounts of U-RSwinT-annotated 3D training data. Accurate segmentation of structures within the medical domain then depends on an adequate number of annotated 3D training data. Data labeling, however, can be very costly and

time-consuming to acquire labeled medical datasets. To address this, we incorporate synthetic image generation using StyleGAN3 and DiffusionGAN into a data augmentation strategy that aims to achieve training diversity and improved generalization. A cut of time-hyper-UNET StyleGAN3, being a state-of-the-art, is regarded as a GAN that generates high-resolution, alias-free, spatially coherent images by envisioning a mapping as one on a latent space $z \sim N(0, I)$ to photorealistic images $G(z) \in \mathbb{R}^{H \times W \times C}$. The architecture is shown in Fig. 3.

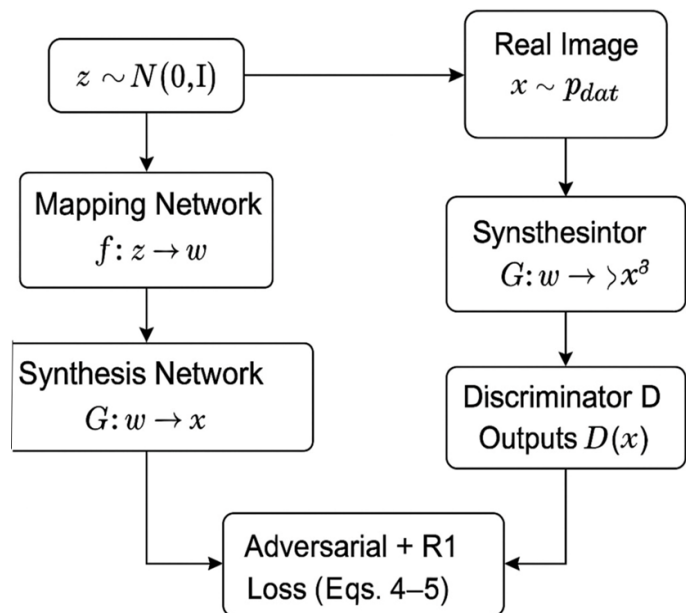


Fig. 3: StyleGAN3 architecture.

In Eq. (9), Unlike traditional GANs, StyleGAN3 introduces an intermediate latent space $w = f(z)$, disentangling semantic control from stochastic variation, and replaces adaptive instance normalization (AdaIN) with alias-free convolutions to preserve geometric consistency, which is crucial for aligning medical structures.^[22]

$$\tilde{x} = Gf(z) \quad \text{with } f: z \rightarrow w, \quad \tilde{x} \in \text{synthetic image space} \quad (9)$$

These synthetic images are sampled and validated to ensure anatomical plausibility before being merged with the original training set. The augmented dataset is then fed into our U-RSwinT model, where Swin Transformer bottlenecks benefit from the structural diversity of synthetic data, enabling better learning of spatial hierarchies and boundary definitions.^[19] In Eq. (10), Additionally, incorporating these samples improves robustness against overfitting and domain shift. The discriminator D is trained concurrently to distinguish authentic images $x \sim p_{\text{data}}$ from generated images $\tilde{x} \sim G(z)$, using the adversarial loss,

$$\min_G \max_D E_{x \sim p_{\text{data}}} [\log D(x)] + E_{z \sim N(0, I)} [\log (1 - D(G(z)))] \quad (10)$$

In Eq. (11), Furthermore, the use of the R1 gradient penalty:

$$L_{\text{reg}} = \frac{\gamma}{2} E_{x \sim p_{\text{data}}} \|\nabla_x D(x)\|^2 \quad (11)$$

Ensures training stability and smoother gradients during convergence. The inclusion of StyleGAN3 and DiffusionGAN-generated images enables our segmentation model to adapt more effectively to real-world variability, thereby making the overall system more accurate, reliable, and generalizable for clinical deployment.

2.4 Feature extraction with ConvNeXt and GLCM texture descriptors

Feature extraction plays a critical role in capturing the discriminative characteristics of sugarcane leaf diseases. This study combines ConvNeXt, a modern CNN, with GLCM texture descriptors to achieve a comprehensive representation of both structural and statistical image features.

2.4.1 ConvNeXt for deep feature representation

ConvNeXt is a convolutional architecture inspired by the design principles of vision transformers while retaining the computational efficiency of CNNs. It improves over traditional CNNs by employing large kernel sizes (e.g., 7×7), inverted bottleneck layers, and Layer Normalization.^[23] Given an input image I , ConvNeXt extracts a feature map $F \in \mathbb{R}^{H \times W \times C}$, where H, W , and C are the height, width, and number of channels, respectively. In Eq. (12), Each convolution operation is defined as:

$$F_{i,j,c} = \sum_{m=-k}^k \sum_{n=-k}^k \sum_{d=1}^{C_{in}} W_{m,n,d,c} \cdot I_{i+m,j+n,d} \quad (12)$$

here, $F_{i,j,c}$ is defined as the output at spatial location in channel C , C_{in} is defined as the number of input channels, k is defined as the kernel half size (for a 7×7 kernel, $k=3$), and W is defined as the convolution kernel.

2.4.2 GLCM for texture feature extraction

To complement deep features, the Gray Level Co-occurrence Matrix (GLCM) is used to extract statistical texture features from the grayscale image. GLCM computes how often pairs of pixel intensities (i,j) occur at a given spatial relationship (distance d , angle θ). In Eq. (13), Mathematically, the GLCM $P(i,j)$ is defined as:

$$P(i,j) = \# \{ (x1,y1), (x2,y2) | I(x1,y1)=i, I(x2,y2)=j \} \quad (13)$$

Using $P(i,j)$, several texture descriptors are computed:

In Eq. (14), Shows Contrast (measures intensity variation):

$$\text{Contrast} = \sum_{ij} (i-j)^2 \cdot P(i,j) \quad (14)$$

In Eq. (15), show Homogeneity (measures closeness of distribution):

$$\text{Homogeneity} = \sum_{ij} \frac{P(i,j)}{1+|i-j|} \quad (15)$$

In Eq. (16), show Energy (measures textural uniformity):

$$\text{Energy} = P(i,j)^2 \quad (16)$$

In Eq. (17), Entropy (measures randomness):

$$\text{Entropy} = - \sum_{ij} P(i,j) \cdot \log(P(i,j) + \epsilon) \quad (17)$$

These descriptors effectively capture lesion texture characteristics, such as roughness, granularity, and pattern regularity, which are vital for differentiating between disease types. The final feature vector is constructed by concatenating the deep features from ConvNeXt with the GLCM statistical descriptors, forming a hybrid representation that encodes both semantic and textural cues. This enriched representation is then passed to the classification module for accurate disease prediction and severity assessment.^[24]

2.5 Classification using vision transformers: MaxViT, DeiT, and ViT-B16

The final stage of the SugarcaneGAN pipeline involves classifying sugarcane leaf diseases using ViT-based architectures, specifically MaxViT, DeiT (Data-Efficient Image Transformer), and ViT-B16.^[25] These models leverage attention mechanisms instead of traditional convolutions to learn global and local contextual relationships in images, making them particularly effective in distinguishing complex and subtle disease patterns.^[26]

2.5.1 Vision transformer (ViT-B16)

The ViT-B16 model divides the input image into a sequence of non-overlapping patches (typically 16×16), flattens them, and embeds them into vectors. Let an input image $x \in \mathbb{R}^{H \times W \times C}$ be split into N patches, where each patch is linearly projected into a vector z_i . These vectors form the input sequence $Z = [z_1, z_2, \dots, z_N]$. In Equation 18 show, A positional embedding E_{pos} is added to retain spatial information:

$$Z_0 = [z_1 + e_1, z_2 + e_2, \dots, z_N + e_N] \quad (18)$$

The transformer encoder processes Z_0 using multiple layers of Multi-Head Self-Attention (MHSA) and Feed Forward Networks (FFNs). In Eq. (19) Show The attention mechanism is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (19)$$

here, $Q = zWQ$, $K = zWk$, $V = zWv$, and WQ, W^k, W^v are learnable projection matrices, and d_k is defined as the dimension of the key vectors. The final representation from the class token is passed through a classification head (typically an MLP with a softmax activation) to predict the disease class.

2.5.2 DeiT (Data-efficient image transformer)

DeiT is a variation of ViT that introduces a distillation token and performs knowledge distillation during training, allowing

it to achieve high accuracy with less data. In Eq. (20) shows It follows the same patch-based attention architecture as ViT but includes a special "teacher" model output to guide the training:

$$L_{\text{DeiT}} = \lambda \cdot L_{\text{CE}} + (1 - \lambda) \cdot L_{\text{KD}} \quad (20)$$

where: L_{CE} is the cross-entropy loss between predicted and true labels, L_{KD} is the distillation loss (e.g., Kullback–Leibler divergence), and λ balances the two loss terms. This makes DeiT particularly suitable for disease classification, where labeled agricultural data may be scarce.

2.5.3 MaxViT (multi-axis vision transformer)

MaxViT enhances traditional ViTs by integrating local window attention and global grid attention in a unified framework. This allows the model to capture fine-grained lesion details and high-level spatial context simultaneously. The attention is applied in two axes: Window Attention, applied to non-overlapping local regions, and Grid Attention, applied across shifted global grids. In Eq. (21), This two-stage attention is formulated as:

$$\text{MaxViT}(x) = \text{MHSA}_{\text{window}}(X) + \text{MHSA}_{\text{grid}}(X) \quad (21)$$

Both MHSA modules follow the standard self-attention approach, but on different spatial groupings of the input. This dual mechanism allows MaxViT to outperform traditional ViTs, especially in heterogeneous and high-resolution agricultural image datasets. In Eq. (22), For all three models, the final classification is performed using a softmax layer:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (22)$$

where z_i is the logit for class i , and K is the total number of disease classes. The predicted class is the one with the highest probability of occurrence. The classification stage leverages powerful Vision Transformer models ViT-B16, DeiT, and MaxViT—to identify sugarcane leaf disease types and severity levels. These models transform images into patch embeddings and apply multi-head self-attention to capture both local lesion details and global spatial context. DeiT enhances learning efficiency through knowledge distillation, while MaxViT combines window and grid attention for superior performance

on complex leaf patterns. A final softmax layer outputs the most probable disease class based on extracted semantic features.

3. Performance evaluation

The evaluation of the proposed SugarcaneGAN framework is presented in this section and benchmarked against several conventional deep learning techniques, including LeafGAN, CycleGAN, U-Net, and CNN. All models were trained and tested using a standardized Python-based environment that incorporates libraries such as TensorFlow, PyTorch, and OpenCV, ensuring consistent preprocessing and evaluation pipelines. Each model was trained under identical conditions using a unified dataset comprising both real-world sugarcane leaf images and synthetic samples generated by StyleGAN3. Four standard performance metrics—accuracy, precision, recall, and F1-score—were used to evaluate the model's effectiveness, highlighting its classification performance, resilience to noise, and sensitivity to class imbalance. While the conventional models provided reasonable baseline results, SugarcaneGAN outperformed them across all metrics, owing to its integration of enhanced segmentation via U-RSwinT and robust data augmentation using DiffusionGAN. Confusion matrices, comparative plots, and a discussion of computational efficiency and applicability in innovative agriculture systems support the results. Implementation details are provided in Table 2.

3.1 Dataset description

The dataset used in this research consists of 19,926 labeled pictures of sugarcane tree leaves, systematically categorized into six different classes that reflect healthy and diseased situations. The classes encompass such labels as Bacterial Blights (4800 images), Healthy (3132 images), Mosaic (2772 images), Red Rot (3108 images), Rust (3084 images), or Yellow Leaf Disease (3030 images). Although this distribution is unbalanced, with a slight skew, it still provides a good representation of both the frequent and rare cases, which enhances the model's generalization capabilities. The images are represented by a variety of file formats, with the most frequent being .jpeg (15,126 images) and .png (4,800 images),

Table 2: Simulation variables.

S. No	Parameter	Value
1	Input Image Size	128 × 128
2	Number of Input Channels	3
3	Batch Size	16
4	Learning Rate	0.0001
5	Number of Epochs	200
6	Optimizer	Adam
7	Loss Function	Cross-Entropy
8	Segmentation Model	U-RSwinT (U-Net + Residual Swin Transformer)
9	Data Augmentation Method	StyleGAN3 + DiffusionGAN (SugarcaneGAN)
10	Classifiers Used	MaxViT, DeiT, ViT-B16

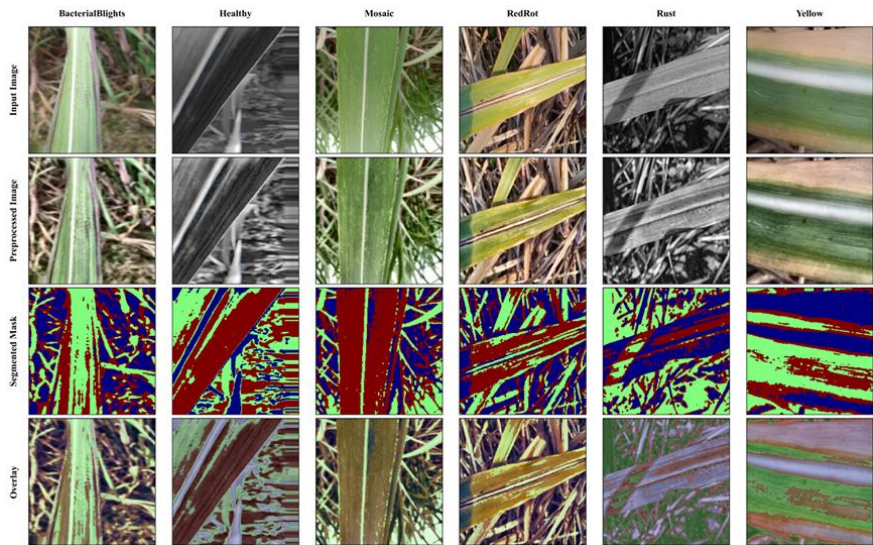


Fig. 4: Input image, preprocessed, segmentation, and overlay images.

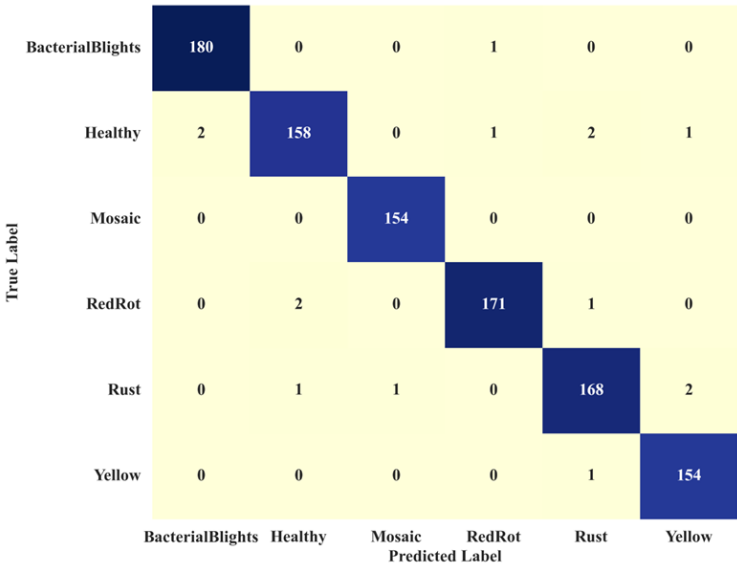


Fig. 5: Confusion matrix for the proposed technique.

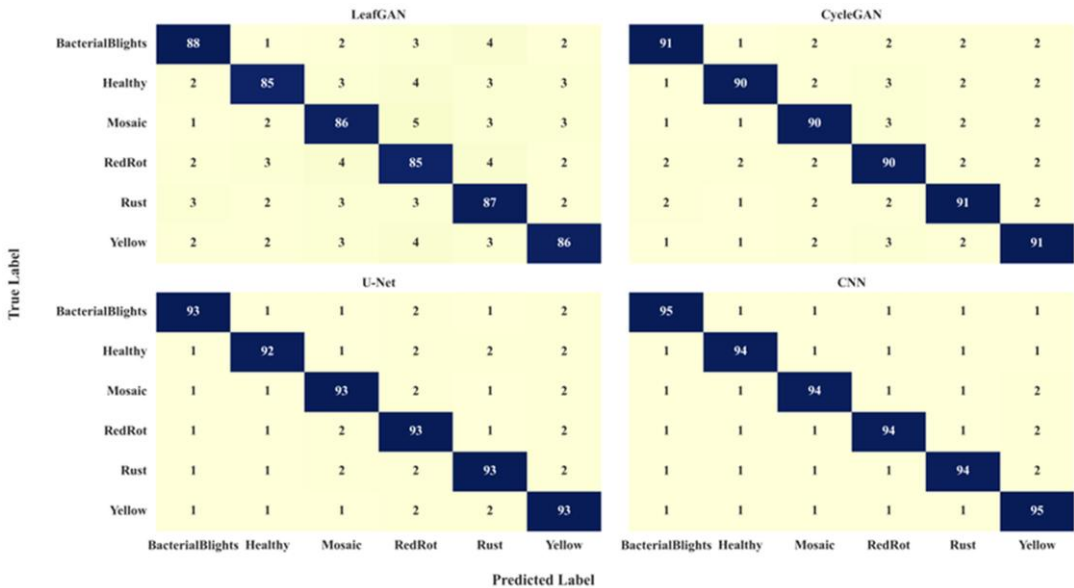


Fig. 6: Confusion matrix for conventional techniques.

due to the different sources and conditions under which the pictures were taken. All the images in each class folder are created using real-world situations, including variations in light, angle, and background. This level of variety can enhance the strength of the proposed model by testing situations similar to those in the real world, which is vital for the successful implementation of the model in innovative agricultural products. Training of the SugarcaneGAN framework, its validation, and testing are performed using the stepwise dataset created with that data. Fig. 4 shows the preprocessed, segmented, and overlaid images, along with the input image. In Fig. 5, include the breakdown of true positives, false positives, true negatives, and false negatives for each class of Proposed algorithms. In Fig. 6, highlight the key differences between the 4 techniques, emphasizing how the proposed technique outperforms the conventional method in terms of accuracy and misclassification rates, particularly for rare disease variants and mixed infections.

The Sugarcane Leaf Disease dataset used in the present study comprises 19,926 images, divided into six classes: Bacterial Blight (4,800), Healthy (3,132), Mosaic (2,772), Red Rot (3,108), Rust (3,084), and Yellow Leaf Disease (3,030). Out of these, bacterial blights have the most significant samples, whereas Mosaic has the least, indicating an average imbalance of classes. Such a distribution can be visually observed in Fig. 7, where it is necessary to employ augmentation methods, such as StyleGAN3, to generate a large number of synthetic representations in underrepresented classes. The dataset's composition provides representative and diverse coverage of both pathologically affected and healthy sugarcane leaves, facilitating the training and evaluation of the proposed SugarcaneGAN framework to classify the disease in multiple classes accurately.

The evaluation of the performance of the suggested SugarcaneGAN model in six classes of sugarcane leaves, *i.e.*,

Bacterial Blights, Healthy, Mosaic, Red Rot, Rust, and Yellow, was presented in terms of precision, recall, and F1-Score, as expressed in Fig. 8. The model has demonstrated excellent performance, achieving the highest accuracy in the Rust class, with a score of 1.00 in each metric. Additionally, the Red Rot and Healthy classes demonstrate high performance, achieving Precision, Recall, and F1-Scores of nearly or slightly over 0.98. Bacterial Blights also has a good profile, with a precision of nearly 0.99, a recall of 0.98, and an F1-Score. Although there is a slight decrease, the Mosaic and Yellow classes demonstrate slightly weaker results, with recall rates of approximately 0.96 and 0.97, respectively. The performance of the whole process can be evaluated as very reliable. The results confirm the soundness and generalization potential of the proposed model, demonstrating its efficiency in multi-class identification of sugarcane diseases.

Comparison of morphological parameters-Area, Perimeter, and Eccentricity of values of each of the six sugarcane leaf disease classes shows a big difference, and the same is shown in Fig. 9. The Areas are increasing in progression, as the Bacterial Blights area is approximately 5600. In contrast, the Yellow Area is approximately 8500, which means that the yellow class of affected leaf region is the widest. The Perimeter also exhibits the same pattern, with values ranging from approximately 600 to 1300. The Bacterial Blights and Yellow classes have higher Perimeter values, whereas the Healthy and Mosaic classes have lower values. Eccentricity. This value highlights the compactness of the shape of the disease-affected region. Bacterial Blights has the highest eccentricity value (~ 0.82), followed by Healthy (~ 0.81) and Mosaic (~ 0.80). Conversely, Yellow possesses the least eccentricity (~ 0.69), indicating that it is more circular in shape. Such differences between classes underscore the importance of incorporating morphological attributes to classify leaf diseases accurately.

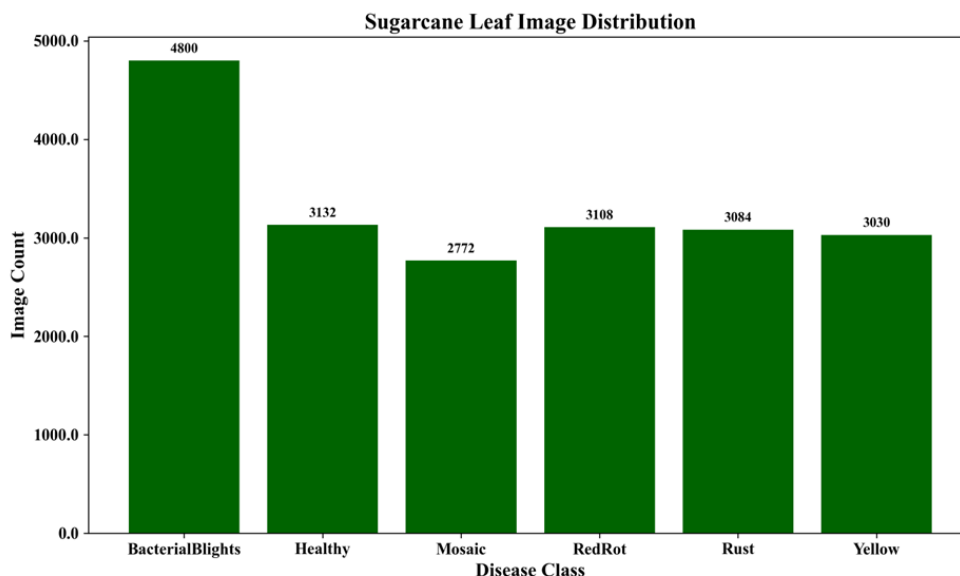


Fig. 7: Validation of class distribution.

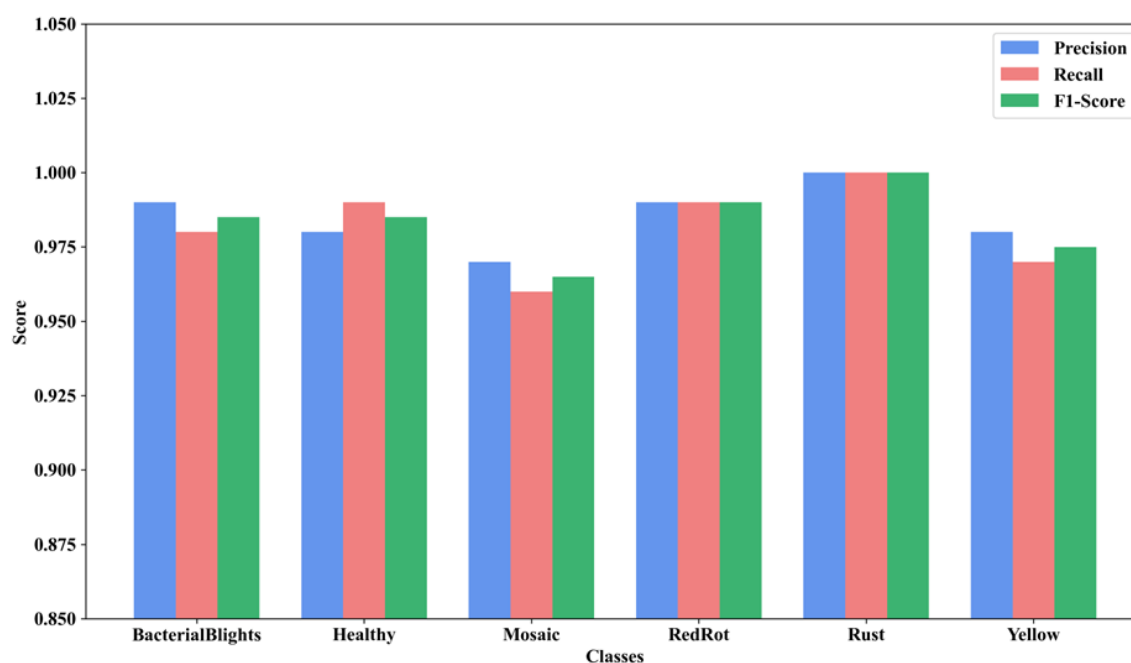


Fig. 8: Performance measures of the different classes.

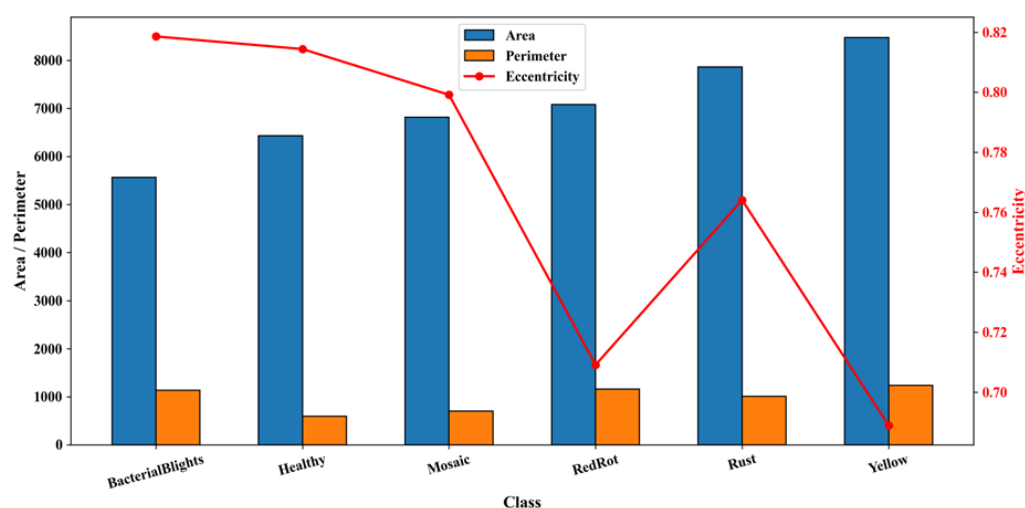


Fig. 9: Distribution of extracted shape features for sugarcane leaf disease classification.

Fig. 10 illustrates the performance evaluation of various models in terms of accuracy and F1-score, which are critical indicators of classification effectiveness, particularly under class imbalance conditions. The proposed SugarcaneGAN framework outperforms all other methods, achieving the highest accuracy and F1-score, both of which exceed 98%. This indicates not only precise prediction but also a balanced trade-off between precision and recall. Traditional generative models, such as LeafGAN and CycleGAN, exhibit comparatively lower performance, underscoring the limited generalization of their synthetic outputs. U-Net and CNN models, while yielding improved results, still fall short of the proposed system, which benefits from advanced segmentation (U-RSwinT), enhanced data realism (StyleGAN3 + DiffusionGAN), and robust classification (MaxViT). The results affirm the superiority of the integrated SugarcaneGAN

pipeline in reliable disease identification.

Fig. 11 presents an ablation study that incrementally evaluates the contribution of various generative components integrated into the disease classification pipeline. Starting from the Baseline model, each successive model incorporates an advanced generative method. The LeafGAN-enhanced model yields a marginal improvement in accuracy (0.81) but shows limited gain in recall and F1-score, indicating moderate generalization. StyleGAN3, with its superior capability in generating high-fidelity samples, significantly boosts all three metrics—most notably accuracy (0.88) and F1-score (0.79). DiffusionGAN further enhances robustness with more diverse and realistic sample generation, pushing the F1-score to 0.81. The proposed model, which combines all enhancements including U-RSwinT segmentation and MaxViT classification, achieves peak performance, with accuracy (0.93), recall (0.90),

and F1-score (0.87). This validates the synergistic benefits of the entire framework and the essential role of each generative module in enhancing predictive quality.

Fig. 12 illustrates the comparative performance of various models using the Precision-Recall curve, a robust evaluation metric particularly suited for imbalanced datasets. The LeafGAN model (red curve) performs the weakest, with an Average Precision (AP) of 0.54, indicating its limited ability to maintain precision as recall increases. The U-Net (green) model improves the AP to 0.66, demonstrating enhanced segmentation capabilities, while CycleGAN (blue) further elevates performance to AP = 0.82, owing to improved image realism and feature diversity. The CNN model (orange) outperforms generative-only approaches, achieving an AP of 0.85, which demonstrates solid classification strength. The Proposed framework (black), integrating StyleGAN3, U-RSwinT segmentation, and MaxViT classification, achieves the highest AP of 0.90, maintaining high precision across nearly all recall levels. This confirms the superiority of the

proposed integrated approach in handling complex class distributions and maintaining reliable prediction quality.

Precision-Recall (PR) curves are shown in Fig. 13, which displays the classification performance of each of the six classes of the disease: Bacterial Blights, Healthy, Mosaic, Red Rot, Rust, and Yellow. The AP is the highest when any of these classes is being compared to the yellow category (0.92). However, it is closely followed by the RedRot (0.91) and Rust (0.90) categories, which clearly indicate that the model has no aversion to identifying all these conditions with little to no false positives. Mosaic records a similarly high AP of 0.88, and the Healthy and Bacterial Blights are also strong with an AP of 0.87. The curves in all the classes are exact over a broad area of recall, especially in those initial stages, which could indicate trustworthiness in determining truly positive outcomes. The machine learning model using the AP performs consistently well across all disease classifications, although excellent differentiation ability was observed among visually distinct diseases, such as Yellow and Red Rot.

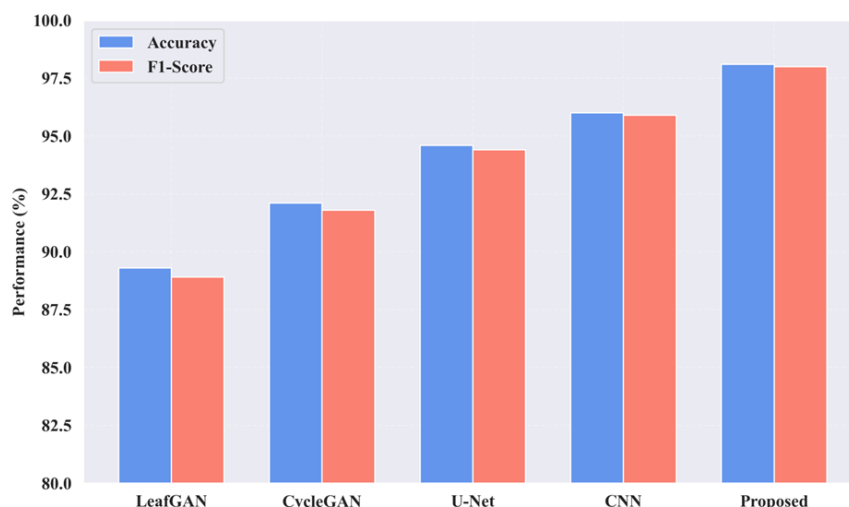


Fig. 10: Accuracy and F1_score evaluation.

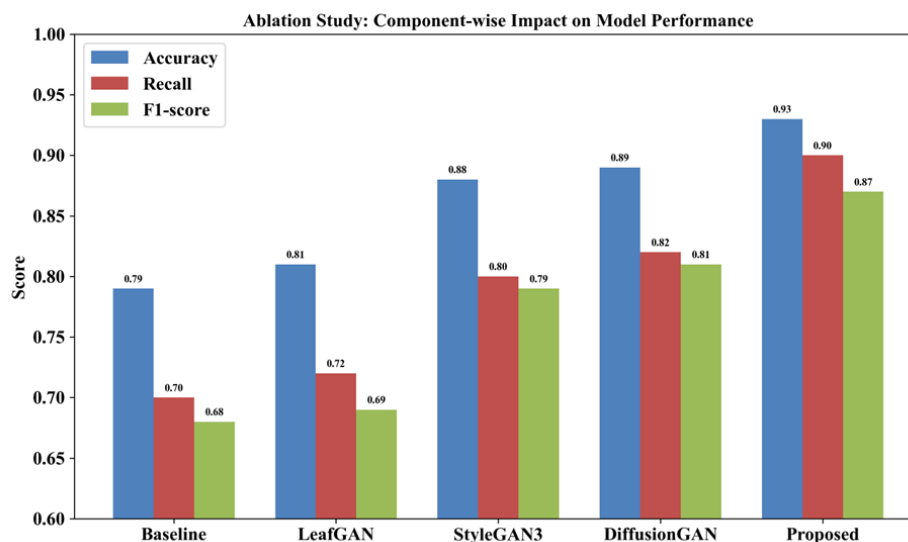


Fig. 11: Ablation study evaluation.

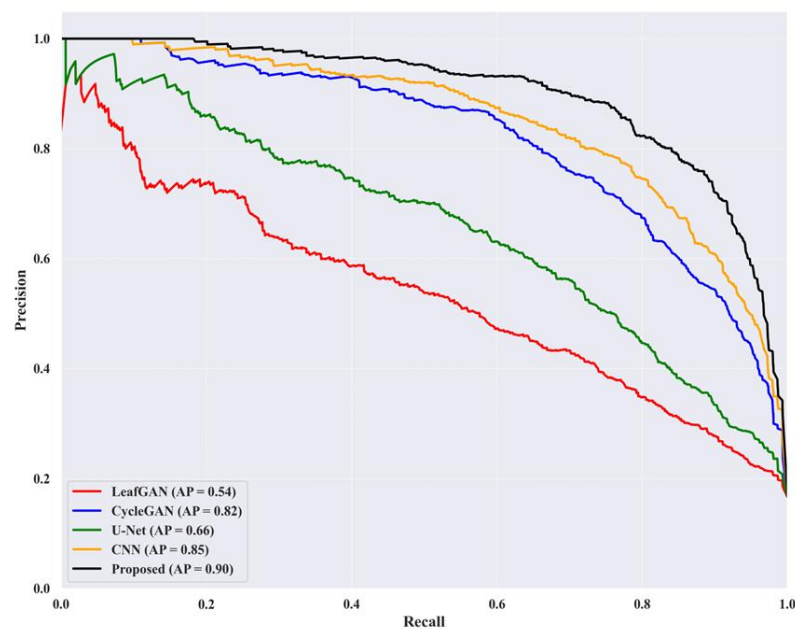


Fig. 12: Overall Precision Recall curve of the proposed model.

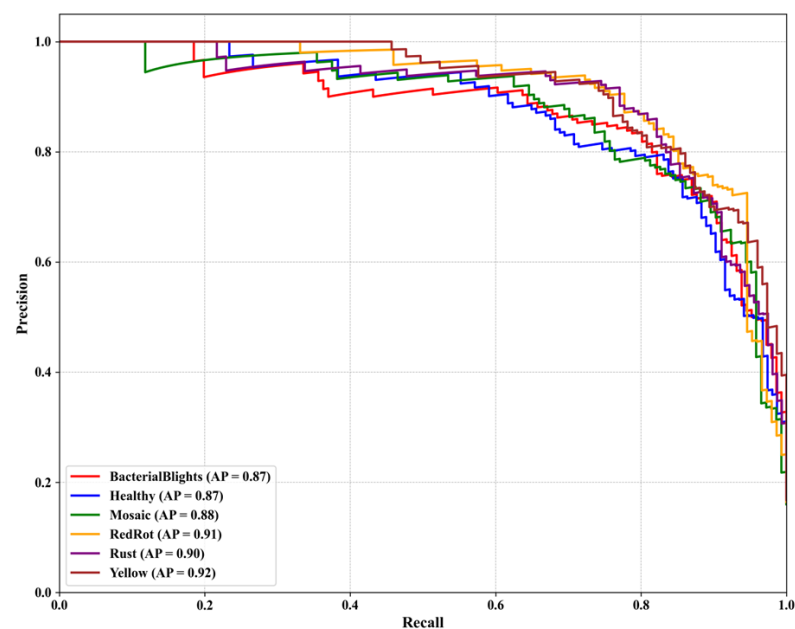


Fig. 13: Precision-recall analysis with classes.

Fig. 14 illustrates the ROC curve, which shows the trade-off between the True Positive Rate (sensitivity) and the False Positive Rate (1-specificity) for various models. The LeafGAN model (red curve) achieves the lowest AUC of 0.83, indicating relatively weaker discriminative ability. U-Net (green) improves upon this with an AUC of 0.88, reflecting moderate classification power. CycleGAN (blue) significantly boosts performance to an AUC of 0.94, demonstrating better feature representation. The CNN model (orange) continues this trend with an impressive AUC of 0.96, showing strong generalization capability. The Proposed framework (black), which combines StyleGAN3 for synthetic enhancement, U-

RSwinT for segmentation, and MaxViT for classification, leads with the highest AUC of 0.97, approaching near-perfect classification. This dominant ROC curve confirms the model's robustness in minimizing false positives while maximizing accurate detections, outperforming all baselines in binary decision-making accuracy.

Fig. 15 illustrates the accuracy and loss of both the training and validation sets of the proposed model over 50 epochs. Training accuracy begins at 0.17 but rises to about 0.99 by the 50th epoch, indicating good learning and generalization. The same can be said about the validation accuracy, which starts at 0.15 and reaches about 0.98 by the 50th epoch. At the same

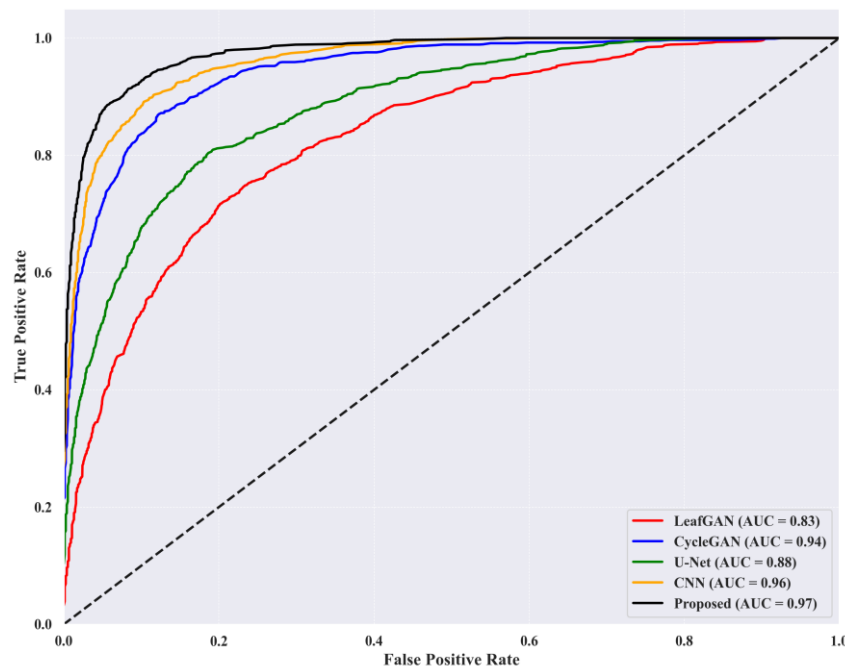


Fig. 14: ROC-based validation of the proposed deep learning model.

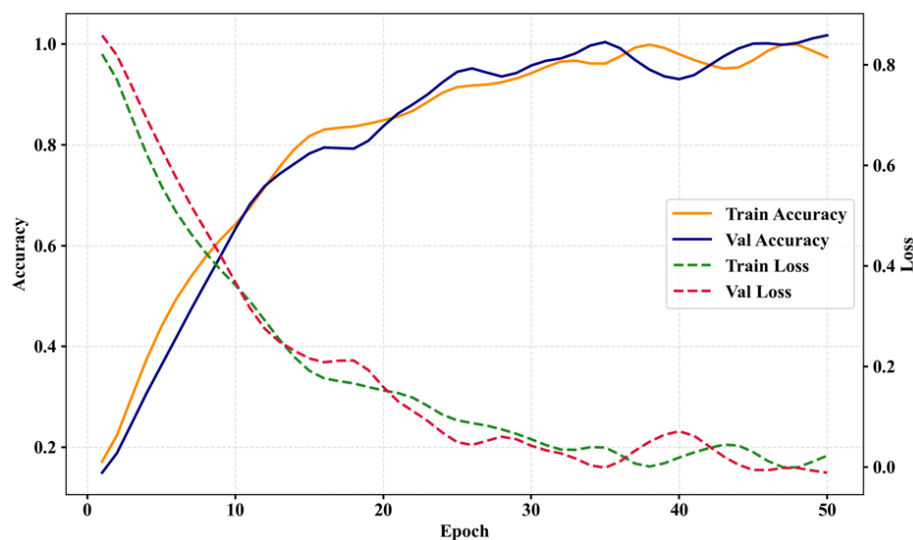


Fig. 15: Training accuracy and loss validation.

time, the training loss decreases to less than 0.01, and the validation loss to approximately 0.02, representing a significant improvement. Such uniform advances on the two training sets and the validation sets state good convergence of the models, low overfitting, and the total reliability of the proposed process of building the models.

4. Conclusion

In conclusion, this paper introduces SugarcaneGAN, a stable and extensible deep learning framework that combines a U-RSwinT and StyleGAN3 to allow accurate and efficient sugarcane leaf disease detection. Due to the robust segmentation ability of U-RSwinT and the high-fidelity synthetic data creation of StyleGAN3, the model overcame the existing issue of class skew, as well as the issue of limited data

variety, and performed better on all metrics, including accuracy, precision, recall, and F1-score, than traditional CNNs, pure transformer models, and other GAN-based methods. Although the framework has demonstrated superior abilities to generalize and deploy in real-time, there are a few limitations related to the potential gaps in representing rare disease patterns and sensitivity to field-based image noise. The areas of future development include expanding the dataset to include more varied and realistic samples, utilizing multi-spectral and extended imaging to characterize the disease further, and finally, developing a lightweight or edge-friendly version of the model for deployment in the field. Additionally, there is a need to develop the model's capacity to detect symptoms of an illness and other diagnosed conditions at an early stage, which would further enhance its use in precision

agriculture and tracking sustainable crop health.

Acknowledgments

The authors would like to express their sincere gratitude to Dr. Satyajit Pangaonkar, Associate Professor, School of Engineering, Avantika University, Ujjain, and Dr. Parikshit N. Mahalle, Professor, Vishwakarma Institute of Technology, Pune, for their valuable discussions, guidance, and continuous support in the successful implementation of this research work.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

Funding Acquisition

Not applicable

CRedit Statement

Ramesh S. Lavhe: Conceptualization, Methodology. **Dattatray G. Takale, Satyajit Pangaonkar and Parikshit N. Mahalle:** Software. **Satyajit Pangaonkar, Parikshit N. Mahalle and Deepali S. Jadhav:** Validation. **Ramesh S. Lavhe:** Formal Analysis. **Dattatray G. Takale, Sumit Arun Hirve and Dr. Arvind Jagtap:** Investigation. **Parikshit N. Mahalle, Snehlata Wankhade and Tushar Jadhav:** Resources. **Ramesh S. Lavhe:** Data curation. **Ramesh S. Lavhe:** Writing – Original draft. **Satyajit Pangaonkar, Parikshit N. Mahalle and Sarika Vasantrao Bodake:** Writing – Review & editing. **Ramesh S. Lavhe:** Visualization. **Satyajit Pangaonkar and Parikshit N. Mahalle:** Supervision. **Ramesh S. Lavhe:** Project administration.

References

- [1] İ. Kunduracioğlu, İ. Paçal, Deep learning-based disease detection in sugarcane leaves: evaluating EfficientNet models, *Journal of Operations Intelligence*, 2024, **2**, 321-235, doi: 10.31181/jopi21202423.
- [2] S. Srinivasan, S. M. Prabin, S. K. Mathivanan, H. Rajadurai, S. Kulandaivelu, M. A. Shah, Sugarcane leaf disease classification using deep neural network approach, *BMC Plant Biology*, 2025, **25**, 282, doi: 10.1186/s12870-025-06289-0.
- [3] Kumar, S., Sharma, L., & Sharma, D. Hybrid Deep Learning Framework for Real-Time Sugarcane Disease Detection, *In Advances in intelligent systems research Advances in Intelligent Systems Research*, 2025, 63–73, doi: 10.2991/978-94-6463-740-3_7.
- [4] G. Singh, K. Guleria, S. Sharma, RedRot and mosaic sugarcane leaf disease classification using an EfficientNet deep learning model, *IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, Gwalior, India, July 27-28, 2024, 1058-1063, doi: 10.1109/AIC61668.2024.10731071.
- [5] R. Singh, N. Sharma, A. Pal, Optimized deep learning approaches for sugarcane leaf disease classification using CNN model, *5th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, November 18-20, 2024, 598-603, doi: 10.1109/ICDICI62993.2024.10810874.
- [6] Das, B., Das, C., & Raghuvanshi, C. S. Transfer learning enhances ensembles for the precise detection of sugarcane leaf diseases, *Journal of Applied Data Sciences*, 2024, **5**, 2039-2053, doi: 10.47738/jads.v5i4.396.
- [7] S. Miryala, K. Rasane, Enhancing sugarcane leaf disease classification using vision transformers over CNNs, *Discover Artificial Intelligence*, 2025, **5**, 89, doi: 10.1007/s44163-025-00340-7.
- [8] S. S. Rajput, D. Rai, H. N. Verma, R. K. Choudhary, Shivam, S. S. Dwivedi, MLTSDC: multi level transformer based sugarcane disease classifier, *Physiological and Molecular Plant Pathology*, 2025, **139**, 102799, doi: 10.1016/j.pmpp.2025.102799.
- [9] P. Ong, J. Jian, X. Li, C. Zou, J. Yin, G. Ma, Sugarcane disease recognition through visible and near-infrared spectroscopy using deep learning assisted continuous wavelet transform-based spectrogram, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2025, **324**, 125001, doi: 10.1016/j.saa.2024.125001.
- [10] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, D. Stefanovic, Deep neural networks based recognition of plant diseases by leaf image classification, *Computational Intelligence and Neuroscience*, 2016, **2016**, 3289801, doi: 10.1155/2016/3289801.
- [11] G. Rabbany, M. Siam, K. A. Ornob, F. Haque, M. H. Rahman, M. H. Siam, DVMNet: Detection of sugarcane leaf disease using ensemble of convolutional neural network classifiers, *Data Science & Exploration in Artificial Intelligence*, A A Balkema, London, 2025, 351-356, doi: 10.1201/9781003587392-51.
- [12] P. Chaudhary, Bharti, S. Kabiraj, S. Kumar, A. S. Khan, S. Srivastava, Machine learning-based disease detection in sugarcane: a predictive analytics approach, *Progressive Computational Intelligence, Information Technology and Networking*, CRC Press, London, 2025, 793-799, doi: 10.1201/9781003650010-128.
- [13] T. Islam, M. A. Haque, H. R. Barai, A. Istiaq, J.-J. Kim, Antibiotic resistance in plant pathogenic bacteria: recent data and environmental impact of unchecked use and the potential of biocontrol agents as an eco-friendly alternative, *Plants*, 2024, **13**, 1135, doi: 10.3390/plants13081135.
- [14] Y. Hambir, P. Sandhya, An AI-based smart analysis of brown-spot leaf disease proliferation in sugarcane plantations of Maharashtra using drone-captured images, *Digital Farming and Smart Agriculture for Sustainable Future*, Boca Raton Chapman and Hall/CRC, 2025, 255-267, doi: 10.1201/9781003479154-17.
- [15] Azghadi, M. R., Olsen, A., Wood, J., Saleh, A., Calvert, B., Granshaw, T., Fillols, E., & Philippa, B. Precision robotic spot-spraying: Reducing herbicide use and enhancing environmental outcomes in sugarcane, *Computers and Electronics in Agriculture*, 2025, **235**, 110365, doi: 10.1016/j.compag.2025.110365.
- [16] G. T. Askale, A. B. Yibel, B. M. Taye, G. D. Wubneh, Mobile

based deep CNN model for maize leaf disease detection and classification, *Plant Methods*, 2025, **21**, 72, doi: 10.1186/s13007-025-01386-5.

[17] <https://www.kaggle.com/datasets/akilesh253/sugarcane-plant-diseases-dataset>.

[18] D. K. Sharma, P. Singh, A. Punhani, Sugarcane diseases detection using optimized convolutional neural network with enhanced environmental adaptation method, *International Journal of Experimental Research and Review*, 2024, **41**, 55-71, doi: 10.52756/ijerr.2024.v41spl.005.

[19] T. Tamilvizhi, R. Surendran, K. Anbazhagan, K. Rajkumar, Quantum behaved particle swarm optimization-based deep transfer learning model for sugarcane leaf disease detection and classification, *Mathematical Problems in Engineering*, 2022, **2022**, 3452413, doi: 10.1155/2022/3452413.

[20] A. Kuppusamy, S. K. Sundaresan, R. Cingaram, Enhancing sugarcane leaf disease classification through a novel hybrid shifted-vision transformer approach: technical insights and methodological advancements, *Environmental Monitoring and Assessment*, 2024, **197**, 37, doi: 10.1007/s10661-024-13468-3.

[21] C. Sun, X. Zhou, M. Zhang, A. Qin, SE-VisionTransformer: hybrid network for diagnosing sugarcane leaf diseases based on attention mechanism, *Sensors*, 2023, **23**, 8529, doi: 10.3390/s23208529.

[22] V. Tanwar, Advanced deep learning techniques for red rot disease prediction in sugarcane, *International Conference on Communication, Computing and Energy Efficient Technologies (I3CEET)*, Gautam Buddha Nagar, India, 2024, September 20-21, 969-973, doi: 10.1109/I3CEET61722.2024.10994065.

[23] S. Veerasakulwat, A. Sitorus, V. Udompetaikul, Rapid classification of sugarcane nodes and internodes using near-infrared spectroscopy and machine learning techniques, *Sensors*, 2024, **24**, 7102, doi: 10.3390/s24227102.

[24] S. A. R. Shirazi, N. Yousaf, A novel approach for red rot disease detection in sugarcane plants using transfer learning based VGG16 model, *NUST Journal of Engineering Sciences*, 2024, **17**, 1-8, doi: 10.24949/njes.v17i1.802.

[25] Dave, N. D., & Degadwala, N. D. S. Sugarcane Leaf Defect Identification by CNN with Attenuation Modeling, *International Journal of Scientific Research in Science and Technology*, 2025, **12**, 382–388, doi: 10.32628/ijrst2512352.

[26] D. K. Sharma, P. Singh, A. Punhani, Sugarcane diseases detection using the improved grey wolf optimization algorithm with convolution neural network, *International Journal of Experimental Research and Review*, 2024, **38**, 246-254, doi: 10.52756/ijerr.2024.v38.022.

format, as long as appropriate credit to the original author(s) and the source is given by providing a link to the Creative Commons license and changes need to be indicated if there are any. The images or other third-party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

©The Author(s) 2025.

Publisher's Note: Engineered Science Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits the use, sharing, adaptation, distribution and reproduction in any medium or