SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I3P125

Volume 13 Issue 3, 312-323, March 2026

Original Article

Enhanced Smart Crop Health Assessment using Interactive Masked Vision Transformer with Multi-Network Attention Mechanism on UAV Imager

Sharmila G

Department of Computer Applications, MLA Academy of Higher Learning, Bangalore, Karnataka, India.

Corresponding Author : sharmiravi75@gmail.com

Received: 18 December 2025 Revised: 21 January 2026 Accepted: 26 February 2026 Published: 23 March 2026

Abstract *- Soybeans have become one of the most significant oilseed and food crops worldwide. However, soybean crops are susceptible to numerous factors. Damage due to pests, illnesses, and other factors exceeds 20 per cent of the world's manufacturing. The usage of Unmanned Aerial Vehicles (UAVs) in crop fields was found to be a significant tool for identifying disease patches, enabling professionals and agriculturalists to make better decisions. Furthermore, in this context, deep learning (DL) has made important developments in Artificial Intelligence (AI). As a result, several studies have employed DL to solve a wide range of diverse problems. In the agricultural sector, DL has gained significant interest in improving crop productivity. This study introduces an Unmanned Aerial Vehicle-Based Soybean Crop Health Monitoring Using Advanced Deep Learning Architectures (UAVSCHM-DLA) model. The aim is to present an intelligent system that is capable of monitoring and assessing soybean crop health using integrated UAV and leaf images. Initially, Histogram Equalisation (HE) and Bilateral Filtering (BF) methods are applied to perform image pre-processing. For effective feature extraction, a vision transformer with the Interactive Mask Self-Attention (IMViT) method is employed. Finally, multiple neural networks with an attention mechanism (MNet-Attn) method are implemented for classification. The comparison of the UAVSCHM-DLA technique illustrated superior accuracies of 98.20% and 97.01% on the leaf and UAV datasets, respectively.*

Keywords *- Crop Monitoring, Crop Health Assessment, Unmanned Aerial Vehicles, Deep Learning, Interactive Mask Self-Attention, Vision Transformer.*

1. Introduction

Soybean (Glycine max) is Brazil's leading agricultural commodity and a critical component of its trade balance and economy. Despite satisfactory numbers, multiple disorders caused by viruses, bacteria, nematodes, and fungi have significantly affected soybean crops under various conditions [1]. An initial analysis of disorders is very significant for pesticide management in the crop, thereby reducing the ecological effects of agrochemicals and financial losses [2]. Pest control commonly involves decisions based on the soybean plant's growth stage and the level of infestation [3].

Moreover, the highest prices of chemicals related to less environmentally harmful practices lead to greater adoption of agricultural precision. Hence, the usage of UAVs in crop domains is considered a significant tool for classifying patches of disorder, enabling farmers and experts to make further decisions [4]. UAV Remote Sensing (RS) models may be used to obtain higher-resolution crop canopy imagery and are broadly employed in agricultural precision crop trait monitoring [5].

Compared to airborne RS models and satellite-based RS techniques, UAV RS techniques are relatively flexible and inexpensive to operate, and they require less space for takeoff and landing [6]. Currently, UAV RS techniques are widely utilised to collect crop trait data. Numerous UAV-driven approaches have been proposed to monitor several crop traits, including biomass, leaf chlorophyll content, crop height, and Leaf Area Index (LAI) [7].

Recently, Machine Learning (ML) models have been employed to detect several crop traits that depend on RS imaging. DL has seen significant developments in AI and ML [8]. The use of DL for soybean observation has attracted considerable interest. An initial study was mainly dedicated to pest detection and disorder, but soon, uses such as cultivar classification, seed counting, phenotyping, and yield forecasting were investigated [9]. Early research has focused on examining various DL techniques and structures across diverse applications and fields. A significant exception is weed recognition, as machinery from diverse producers can not only identify the weeds but also remove the issues [10].

This study introduces an Unmanned Aerial Vehicle-Based Soybean Crop Health Monitoring Using Advanced Deep Learning Architectures (UAVSCHM-DLA) model. The key contributions are:

- The Histogram Equalisation (HE) and Bilateral Filtering (BF)-based image pre-processing improves contrast, eliminates noise, and conserves edges.
- Furthermore, a vision transformer with the Interactive Mask Self-Attention (IMViT) method is employed for effective feature extraction and to capture both global and local features for enhanced crop health classification.
- Moreover, multiple neural networks with an attention mechanism (MNet-Attn) based classification emphasize the most relevant regions and enhance accuracy.
- Finally, the experimentation of the UAVSCHM-DLA technique is conducted under the MH-SoyaHealthVision dataset.

2. Existing Research on Soybean Crop Health Monitoring using UAV Images

This section briefly reviews prior related works on soybean crop health assessment. Yu et al. [11] presented a fully automatic approach to assess soybean growth uniformity and emergence rate, applicable to both UAV and ground-driven environments. The FEL-YoloV8 technique was developed by improving the feature extractor, enhancing the Feature Fusion Module (FFM), and integrating lightweight model mechanisms (MLMs). Based on the recognition outcomes from the FEL-YoloV8 technique, counts, missing seedling locations, and soybean growth uniformity are assessed. This approach enables fully automated observation of soybean growth uniformity and emergence rates. Aqeel et al. [12] introduced an AI-drone surveillance method that uses the YOLOv12 model to classify crop disorders in real-world settings automatically. The comprehensive pipeline includes intelligent pre-processing, automated image acquisition, and real-time analysis. This method, when combined with conventional manual inspection, reduces diagnosis time from days to minutes while improving consistency. Main innovations comprise an optimised technique structure for resource-limited settings and multispectral disorder pattern detection. Domain research confirms the method's robustness across different weather states and development phases. Zhang et al. [1] proposed DS-SoybeanNet to enhance the efficiency of UAV-driven observation of soybean maturity data. This method could obtain and use either deep or shallow image characteristics. Recently, the widely used DL technique has focused on deep image feature extraction, whereas shallow image characteristics have been overlooked. In [13], an HSDT-TabNet technique is presented for grading soybean FLS under domain states by analysing UAV-driven hyperspectral information. The method uses a two-path parallel feature extractor approach: sparse feature selection is performed by the TabNet path for comprehending fine-tuned local discriminative data. Kabir et al. [14] proposed a new Quadcopter Drone Crop Observation method that addresses the difficulties farmers face in observing and managing their crops. This method uses an advanced quadcopter drone equipped with a higher-resolution camera. A device would take detailed images of crops from many angles, which are then transferred to a ground station.

Wallinger et al. [15] implemented and evaluated cost-effective, smart-connected methods to support crop management. Especially installed nodes with several sensors in various parts of a greenhouse to monitor environmental and lighting conditions and observe plant development. Moreover, the observing method involves a central access node that accumulates sensory information, which can also be accessed remotely for further analysis. Rahman et al. [16] presented ML for identifying crops from UAV-driven multispectral imagery. Conventional mapping and detection methods capture time and involve numerous individual steps and are frequently subjective. The identification utilises Random Forest Classification as the ML model. Murugan et al. [17] reviewed how Neural Network (NN) methods, including Convolutional NNs (CNN), vision transformers with interactive mask self-attention (IMViT), and Multiple NNs with Attention (MNet-Attn), for image-based detection of maize leaf diseases. Tong et al. [18] developed a lightweight, task-adaptive Task-oriented Transformer Network (ToT-Net) based on the RT-DETR method for accurate, real-time detection of crop diseases. Nouman Noor et al. [19] developed a computer-aided approach by using advanced image processing and DL, specifically You Only Look Once version 8 (YOLOv8) with Transfer Learning (TL) methods. Gopinath and Sangeetha [20] presented a spatiotemporal digital twin using Convolutional Long Short-Term Memory (ConvLSTM) and Graph NNs (GNNs) models. Kaur, Priya, and Singh [21] developed a platform using CNNs (CNN1–CNN4) and a hybrid SMOTE-ENN sampling technique for improved classification and lesion-based severity assessment. Parganiha and Verma [22] introduced a system by utilizing Modified Residual U-Net (MRU-Net), CNN, Improved Vision Transformer (IViT), and a stacking ensemble of XGBoost (XGB), Gradient Boosting (GB), and AdaBoost-Decision Tree (AdB-DT). Kang et al. [23] developed a framework by employing CNN and Transformer-based All-Attention Transformer (A2Former) for accurate semantic segmentation and efficient multi-scale spectral–spatial feature extraction. Du et al. [24] presented a unified transformer-based framework, Unified Hyperspectral Transformer with Prototype-guided Token Routing (UniHSFormer-X) model for accurate and interpretable crop classification from hyperspectral images.

Though the existing studies are significant for evaluating crop health, they exhibit limitations due to limited generalization in real-field conditions. Furthermore, several

methods, including YOLOv8, ToT-Net, DS-SoybeanNet, and UniHSFormer-X, encounter difficulty with data imbalance, noise, and environmental variability. Also, a few techniques, such as MRU-Net with stacking, CNN-based banana disease models, and ConvLSTM-GNN digital twins, show computational complexity. Semantic segmentation and transformer-based techniques, namely A2Former and HSDT-TabNet, improve extraction but need massive labelled datasets and intensive preprocessing. Lightweight and interpretable solutions for multi-crop, multi-disease scenarios are also scarce. There exists a research gap in addressing scalable, field-ready, and computationally efficient crop disease detection systems for high accuracy.

3. Proposed Model

This study follows a systematic approach to monitoring and assessing soybean crop health that blends image pre-processing, extraction, and classification. Figure 1 exhibits the workflow of the UAVSCHM-DLA approach.

- Apply HE and BF to enhance contrast and reduce noise in the image.
- Use an IMViT to derive rich features from pre-processed images.
- Implement MNet-Attn to classify soybean plant health by analysing extracted features.

3.1. Image Pre-processing

This is applied to the obtained images, comprising models to reduce noise and enhance quality and contrast. The imaging conditions are predicted using the BF and HE processes [25]. BFs effectively mitigate image noise without sacrificing edges. It leverages an iterative BF to filter noise while conserving fine structures, thereby enhancing denoising and mitigating bias. The calculation of the range filter coefficient relies on the radiometric distance from the pixel's neighbourhood. The response of BF at pixel point x is depicted.

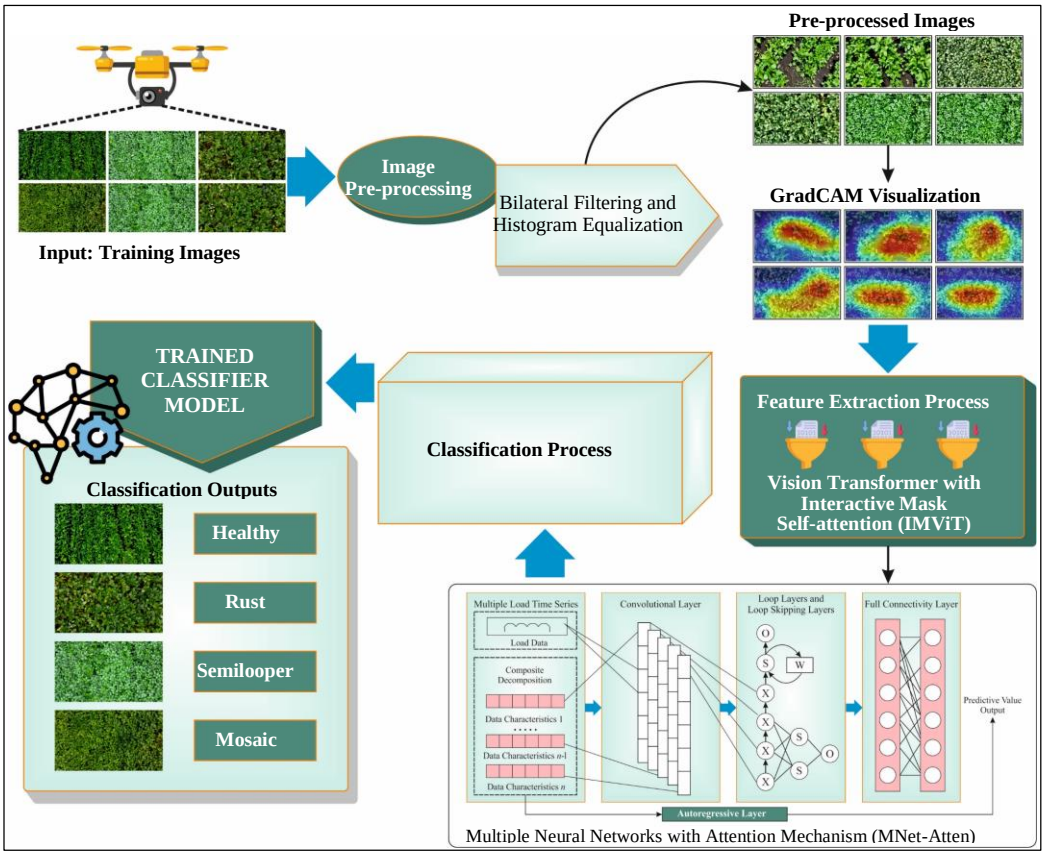

Fig. 1 Workflow of the UAVSCHM-DLA approach

$$\hat{I}(x,y) = \frac{1}{NC} \sum_{y \in N(x)} r_s(x,y) * r_r(x,y) * I(y) \quad (1)$$

$$\mathrm{NC} = \sum_{y \in N(x)} r_s(x,y) * r_r(x,y) \quad (2)$$

Here, r_s and r_r denote radiometric components and the bi-directional filter domain, respectively, although NC denotes the normalisation factor. $N(x)$ represents the neighbourhood of x, and y is the position:

$$r_s(x,y) = \exp\left(\frac{-|x-y|^2}{2\sigma_s^2}\right) \quad (3)$$

$$r_r(x,y) = \exp\left(\frac{-|it_X - it_y|^2}{2\sigma_r^2}\right) \quad (4)$$

Reduced-weight functions r_r and r_s are the outcomes of the radiometric distance between intensities it_x and it_y. N regulates the support of the filter, and σ_s, σ_r, and σ_r determine how smoothly dual weight components decay. The optimum values for σ_s and σ_r are chosen.

HE: This model is utilised to improve the contrast of lower-contrast images. Assume an input image im, which is an m by n integer pixel matrix intensity divided in $[0, L-1]$. L signifies the probable number of intensities among 1 (for double class) and 256. In this instance, h defines the normalised histogram depending on probable intensity:

$$h_n = \frac{No.\ of\ pixels\ with\ intensity\ n}{Overall\ pixels} \tag{5}$$

Here $= 0,1, \dots, L-1$.

$$eq_{(m,n)} = \text{floor}\left(L-1 \sum_{n=0}^{im_{(m,n)}} h_n, \right. \tag{6}$$

Now $floor\ (\cdot)$ signifies the rounded floor value. The vital assumption is that the intensities h and eq are continuous, arbitrary X and Y variables, respectively, with support $[0, L-1]$. Y is denoted.

$$Y = d(X) = (L-1) \times \int_0^X \text{hi}\,(x)dx, \tag{7}$$

Here, d represents the cumulative dispersion function of X multiplied by $(L-1)$, and hi depicts the histogram.

3.2. Feature Extraction

For feature extraction, an IMViT model is used to produce more discriminative, contextually rich representations from pre-processed imagery. The IMViT block aims to extract both global and local visual clues and combine them within each IMViT block [26]. In particular, the IMViT block is designed like a typical transformer encoder layer. Still, it differs in its attention mechanism (AM) and in the insertion of an interaction mechanism into Multi-Head (MH) Self-Attention (SA).

3.2.1. Multi-Receptive AM

MH Attention (MHA) is a standard design in transformer-based models. Conventional MHA designs are identical, limiting their ability to comprehend both local and global information. For instance, the basic ViT uses global attention in every head, whereas Swin uses only local attention. In contrast, IMViT projects dissimilar receptive fields across diverse attention heads to compel the transformer layer to learn global and local information simultaneously. Notably, the receptive areas of 1 × 1, 4 × 4, 7 × 7, and 14 × 14.1 are projected. In the mathematical formulation, set an input feature mapping F_l of the $l-th$ transformer block, successive IMViT blocks are calculated as:

$$z^l = LN(M - MSA(F^{l-1}, M), I) + F^{l-1} \tag{8}$$

$$F^l = LN\big(MLP(z^l)\big) + z^l \tag{9}$$

Here, the input to the initial transformer layer is denoted as $F^0 = F$, $M\text{-}MSA$ denotes a masked attention with dissimilar receptive fields $M = \{1\times1, 4\times4, 7\times7, 14\times14\}$, I represents a learnable vector, $I = \{\alpha, \alpha, \dots, \alpha_n\}$. In particular, the MH SA is still a fully parallel design. A global self-attention like ViT is initially performed, and then produces local masked attention by constructing dissimilar adjacency matrices among tokens. Lastly, different types of masks and global attentions are assigned to multi-heads to achieve local-global self-attention.

3.2.2. MH Interaction Mechanism

Each head computes its output independently, losing the ability to weight the relative importance of global and local data within each transformer block. Specifically, assume the outputs J of H head, *i.e.*, $\hat{J} = (\hat{J}^{(1)}, \hat{J}^{(2)}, \dots, \hat{J}^{(H)})$. The interaction attention mappings $J = (J^{(1)}, J^{(2)}, \dots, J^{(H)})$ is computed by

$$\begin{aligned} \hat{J}^{(1)} &= \cos\,(Q^{(1)}K^{(1)}/\tau) + B, \cdots, J^{(h)} \\ &= \cos\left(\frac{Q^{(h)}K^{(h)}}{\tau}\right) + B. \end{aligned} \tag{10}$$

$$\begin{pmatrix} J^{(1)} \\ J^{(2)} \\ \vdots \\ J^{(h)} \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1H} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2H} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{h1} & \lambda_{h2} & \cdots & \lambda_{HH} \end{pmatrix} \begin{pmatrix} \hat{J}^{(1)} \\ \hat{J}^{(2)} \\ \vdots \\ \hat{J}^{(h)} \end{pmatrix} \tag{11}$$

Here, d denotes a dimension of channels, B represents a learnable bias matrix, and $\Lambda = (\lambda_{ij})_{i=1,j=1}^{i=H,j=H}$ means a learnable parameter. Then, the interaction outputs are computed by the multiplication of attention mappings and values, which is formulated as:

$$\begin{aligned} P^{(1)} = softmax\,J^{(1)}, \dots, P^{(h)} = softmax\,J^{(h)}. \\ O^{(1)} = P^{(1)}V^{(1)},, \cdots, O^{(h)} = P^{(h)}V^{(h)}. \\ O = \left[O^{(1)}, O^{(2)}, \dots, O^{(h)}\right]. \end{aligned} \tag{12}$$

Through an interaction mechanism, the attention mappings of dissimilar heads can cooperate to combine local and global information.

3.2.3. Looping the Receptive Fields Among Multiheads in One Layer

Each head must perceive both global and local information. So, a loopy multi-scale receptive attention is presented, where stated 1×1, 4×4, 7×7, and 14×14 are alternated amongst dissimilar manifold attention heads. The MH SA idea is to replace the masked attention values with a smaller number and keep the rest unchanged. For the circulation mechanism, it is only necessary to insert a dissimilar Prior-adjacency-matrix in each layer according to the law.

3.3. Crop Health Classification using MNet-Attn

For crop health classification, the proposed model employs MNet-Attn to classify soybean crop health conditions using rich feature representations effectively. The MNet-Atten method is developed to extract shorter- and longer-term data dependencies within a non-linear module containing loop-skipping, cyclic, and convolutional layers [27]. The temporal pattern AM was employed to focus on the main series and remove disturbing factors. The linear features were removed using an autoregressive method, and the outcomes were obtained by combining the outputs of the non-linear and linear parts.

3.3.1. Convolution Module

This is a CNN without a pooling layer. It removes the shorter-term local assets and variable dependencies and delivers them to the loop-skip and loop modules. The k^{th} filter achieves the convolutional process on a matrix X_t as exposed below:

$$h_k = R(W_k \times X_t + b_k) \tag{13}$$

Whereas h_k denotes an output feature vector, R means a ReLU activation function, W_k signifies the weight matrix for the kth feature mapping, X_t refers to an input vector, and b_k is the bias vector.

Recurrent and Recurrent-Skip Module: The output of the convolutional component is sent to the loop and loop-jump modules simultaneously. The Bi-GRU method consists of dual layers with similar output but in opposed directions, which can obtain data from forward as well as backward directions, overcoming the defects of conventional GRU in unidirectional data transfer, and can completely develop the time-based features to enhance the rate of data utilisation and a model's accuracy. Bi-GRU is chosen to create the modules of loop skip and loop.

The Hidden Layer (HL) of the loop module at the tth time is computed below:

$$h_t^f = \delta\left(W_{f1}x_t + W_{f2}h_{i-1}^f\right) \tag{14}$$

$$h_t^b = \delta(W_{b1}x_i + W_{b2}h_{t+1}^b) \tag{15}$$

$$h_t = \delta\left(W_1 h_t^f + W_2 h_t^b\right) \tag{16}$$

Whereas x_i denotes an input at tth time, δ refers to a function of sigmoid, h_t^f and h_i^b means an output of positive- and negative-order GRU at tth time, correspondingly, W_{f1} and W_{f2} denote a weight vector of a positive-order GRU, W_{b1} and W_{b2} represent a weight vector of a negative-order GRU model, and W_1 and W_2 means an HL weight vector of positive and negative-order GRU method, correspondingly.

Bi-GRU's bidirectional loop framework takes longer-term temporal relations among information. For ultra-long-term dependencies, Bi-GRU cannot remove effective attributes. To resolve this issue, the loop-skipping module was presented. The loop-skipping module formulation was mentioned below:

$$h_t^f = \delta\left(W_{f1}x_i + W_{f2}h_{t-p}^f\right) \tag{17}$$

$$h_t^b = \delta\left(W_{b1}x_t + W_{b2}h_{t+p}^b\right) \tag{18}$$

Whereas, p signifies the no. of cells in HL.

The outcome of the loop module at the tth time is h_t^R, and the outcome of the loop skip module from $t-p+1$ to the t moment is h_t^S. These outputs are combined via an FC network, and the final output is used for the non-linear part. Its mathematical formulation is computed below.

$$h_t^D = W^R h_t^R + \sum_{t=0}^{p-1} W_i^S \, h_{t-i}^S + b \tag{19}$$

Whereas, h_t^D means an output of the non-linear part at the tth time.

Modelling of Temporal Pattern Attention: To avoid losing the features of high-dimensional data, an AM was developed and widely employed in the model. Traditional AMs relied on calculating correlations at a single time step, thus making data classification difficult.

The time-based Pattern Attention (PA) model extracts features from the HL matrix by utilizing a 1D CNN to capture the essential links between time series and dissimilar attributes.

The input sequence was handled employing BiGRU to attain the hidden features $h_{t-w} - h_t$, $w$here w denotes the length. Describe $H = (h_{t-w}, h_{t-w+1}, \ldots, h_{t-1}\}$ as a hidden feature, H^C means a time-based pattern matrix, and C means the filter. The mapping and computation procedure of the time-based PA model is expressed below:

$$H_{i,j}^C = \sum_{i=0}^{p-1} H_{i,t-w+l} \times C_{j,T-w+l} \tag{20}$$

$$f(H_i^C, h_t) = (H_i^C)^T W_a h_t \tag{21}$$

$$\alpha_i = \delta\left(f(H_i^C, h_t)\right) \tag{22}$$

$$y_{t-1+\Delta} = |v_{h'}(W_h W_t + W_v W_t) \tag{23}$$

Whereas $H_{i,j}^C$ denotes an eigenvalue, α_i denotes an attention weight, T represents a maximum length, f means an evaluation function, δ signifies the function of sigmoid, v_t indicates an attention vector, $y_{t-1+\Delta}$ is forwarded to the softmax classifier for category prediction, Δ denotes the time-step, $W_a, W_{h'}, W_h$, and W_v denotes corresponding variables with dissimilar weight matrices.

Modelling of Autoregressive Module: Owing to the non-linear nature of recurrent and convolutional modules, an autoregressive method is employed as a final method for linear data. Its mathematical formulation is expressed by:

$$h_t^L = \sum_{k=0}^{q^{ar}-1} W_k^{ar} y_{t-k} + b^{ar} \quad (24)$$

Whereas, W_k^{ar} and b^{ar} denote a coefficient of the model, q^{ar} represents a dimension of the input matrix, and h_t^L means a linear part.

The non-linear output part h_t^D and the linear output part h_t^L Weigh the last outcome.

$$\hat{Y}_t = h_t^D + h_t^L \quad (25)$$

Here, $\hat{Y}_t$ is then passed through a softmax layer to obtain class probabilities at time t.

4. Findings and Interpretation

The performance of the UAVSCHM-DLA model is tested on the MH-SoyaHealthVision dataset [28]. Table 1 provides details of the drone's specifications. The image was captured with a DJI FC8482 drone-mounted camera for high-resolution aerial imaging.

Table 1. Details of the drone specification

Device Type	Drone
Camera model	DJI-FC8482
F-stop	f/1.7
Exposure time	1/3000 sec
ISO speed	ISO-110
Focal length	7mm
35mm focal length	24
Modality	s-RGB
Resolution	3840∗2160

The MH-SoyaHealthVision is a wide-ranging dataset developed for incorporating crop health assessment in soybean farming. It integrates UAV-captured images from soybean fields in the Maharashtra region, enabling a holistic model for detecting disease and pest attacks. The leaf image dataset comprises high-resolution visuals of soybean leaves impacted by diseases. Complementing this, the UAV dataset provides large-scale aerial views of soybean fields, capturing rust patterns. The inclusion of UAV technology in this dataset is vital for precision agriculture, as drones enable highly precise, targeted pesticide spraying. The first part comprises the Soybean Leaf Image Dataset, organised into six folders: "Healthy," followed by four folders depicting various diseases, and the last for pest attack. The second part contains the Soybean UAV Image Dataset, classified into four folders: "Healthy," 2 folders depicting diseases, and one folder for pest attack. The leaf dataset comprises 2835 images across six classes, as shown in Table 2. Figure 2 depicts sample images of healthy, rust, semilooper, and mosaic leaves, and Figure 3 shows images of the original, pre-processed, GradCam, and feature maps.

Table 2. Details of the leaf dataset

Leaf Dataset		
Class Name	**Labels**	**No. of Images**
Healthy	Class 1	204
Mosaic	Class 2	707
Rust	Class 3	852
Spectoria Brown Spot	Class 4	268
Frogeye leaf spot	Class 5	222
Caterpillar and Semi-Lopper Pest Attack	Class 6	582
Total		**2835**

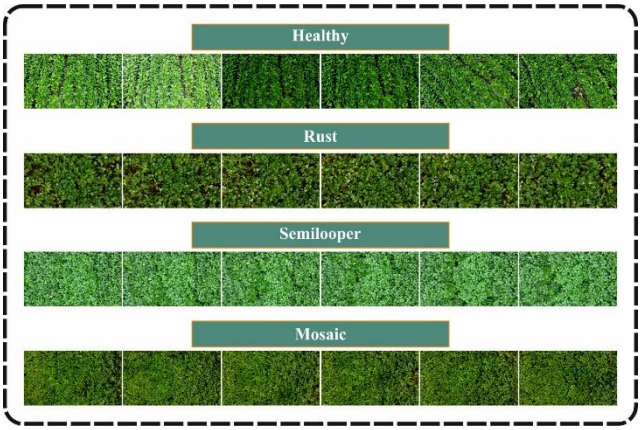

Fig. 2 Sample images of healthy, rust, semilooper, and mosaic leaves

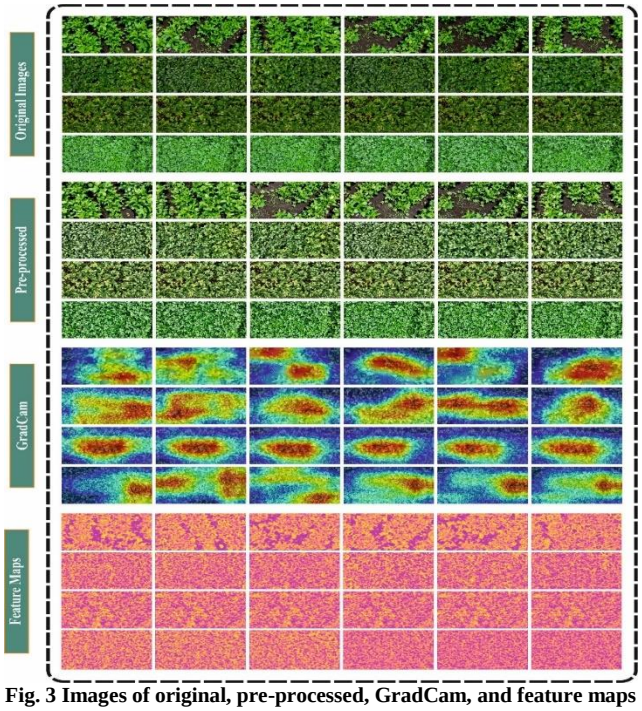

Fig. 3 Images of original, pre-processed, GradCam, and feature maps

Figure 4 signifies the crop health monitoring of the UAVSCHM-DLA approach under 70:30 of the Training Phase (TRAP) and Testing Phase (TESP). Figure 4(a) shows the confusion matrix. Figures 4(b) and 4(c) depict the PR and ROC investigation, showing strong performance.

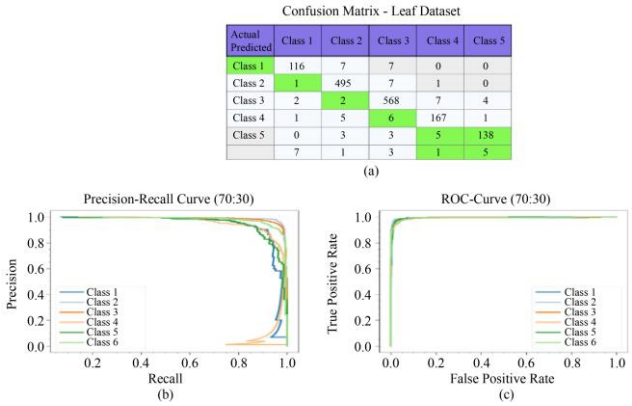

Fig. 4 (a) Confusion matrix, (b) PR, and (c) ROC curves under 70:30 on the leaf dataset.

Table 3 and Figure 5 denote the crop health monitoring of the UAVSCHM-DLA technique under 70:30 on the leaf dataset. The outcome suggests that the UAVSCHM-DLA method appropriately recognised the instances. With 70%TRAP and 30%TESP, the UAVSCHM-DLA method presents an average $accur_y$, $preci_n$, $recal_l$, $F_{measure}$, and AUC_{score} of 98.20%, 93.72%, 92.05%, 92.84%, and 95.46%, and 97.45%, 91.48%, 89.55%, 90.45%, and 93.97%, respectively.

Table 3. Crop health monitoring of the UAVSCHM-DLA technique under 70:30 on the leaf dataset

Class Labels	$Accur_y$	$Preci_n$	$Recal_l$	$F_{measure}$	AUC_{score}
TRAP (70%)					
Class 1	98.34	91.34	84.06	87.55	91.73
Class 2	98.59	96.49	98.02	97.25	98.40
Class 3	97.68	95.62	96.60	96.11	97.37
Class 4	98.29	92.27	89.30	90.76	94.26
Class 5	98.59	93.24	88.46	90.79	93.96
Class 6	97.73	93.35	95.85	94.58	97.04
Average	**98.20**	**93.72**	**92.05**	**92.84**	**95.46**
TESP (30%)					
Class 1	97.88	90.00	81.82	85.71	90.53
Class 2	97.65	94.17	96.04	95.10	97.10
Class 3	96.83	94.38	95.45	94.92	96.45
Class 4	98.12	92.21	87.65	89.87	93.44
Class 5	97.88	88.71	83.33	85.94	91.22
Class 6	96.36	89.39	93.02	91.17	95.11
Average	**97.45**	**91.48**	**89.55**	**90.45**	**93.97**

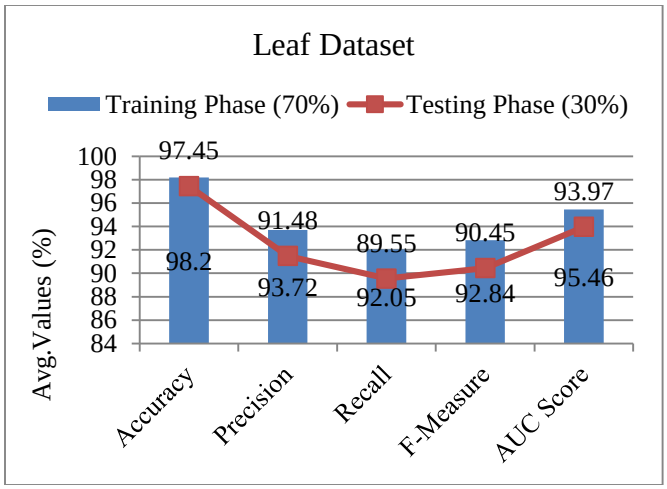

Fig. 5 Crop health monitoring of UAVSCHM-DLA technique under 70:30 on leaf dataset

Figure 6 reveals the Training (TRAN) and Validation (VALD) $accur_y$ of the UAVSCHM-DLA model over 200 epochs. The slow convergence of the model suggests that VALD is superior to TRAN, indicating robust outcomes and stability.

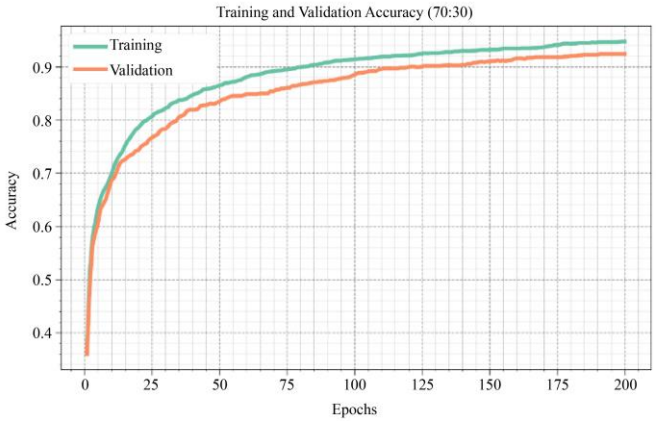

Fig. 6 $Accur_y$ curve of UAVSCHM-DLA model under 70:30 on leaf dataset

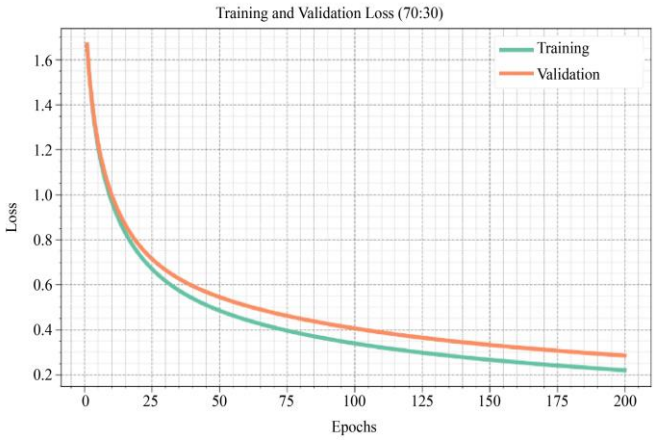

Fig. 7 Loss curve of UAVSCHM-DLA model under 70:30 on leaf dataset

Figure 7 demonstrates the TRAN and VALD loss of the UAVSCHM-DLA method over 200 epochs. The VALD remained lower than the TRAN loss, highlighting better generalisation and no overfitting.

The comparative outputs of the UAVSCHM-DLA method with existing techniques are presented in Table 4 and Figure 8 [29, 30]. The analysis specified that the UAVSCHM-DLA model outperformed an enhanced outcome. Based upon $accur_y$, the UAVSCHM-DLA model has an $accur_y$, $preci_n$, and $F_{measure}$ of 98.20%, 93.72%, and 92.84%, whereas the Resnet, Vggnet, Vit-B, Vit-L, Swin-T, and ED-Swin Transformer models have a lower $accur_y$ of 78.32%, 78.18%, 85.82%, 88.72%, 92.12%, and 94.32%, and $preci_n$ of 74.04%, 74.74%, 85.40%, 87.84%, 92.08%, and 94.56%, and $F_{measure}$ of 82.93%, 83.28%, 90.66%, 92.21%, 94.94%, and 96.52%, respectively.

Table 4. Comparison of outcomes of UAVSCHM-DLA techniques with existing approaches on the leaf dataset

Leaf Dataset				
Methods	$Accur_y$	$Preci_n$	$Recal_l$	$F_{measure}$
Resnet	78.32	74.04	94.24	82.93
Vggnet	78.18	74.74	94.02	83.28
Vit-B	85.82	85.40	96.62	90.66
Vit-L	88.72	87.84	97.04	92.21
Swin-T	92.12	92.08	97.98	94.94
ED-Swin Transformer	94.32	94.56	98.56	96.52
UAVSCHM-DLA	98.20	93.72	92.05	92.84

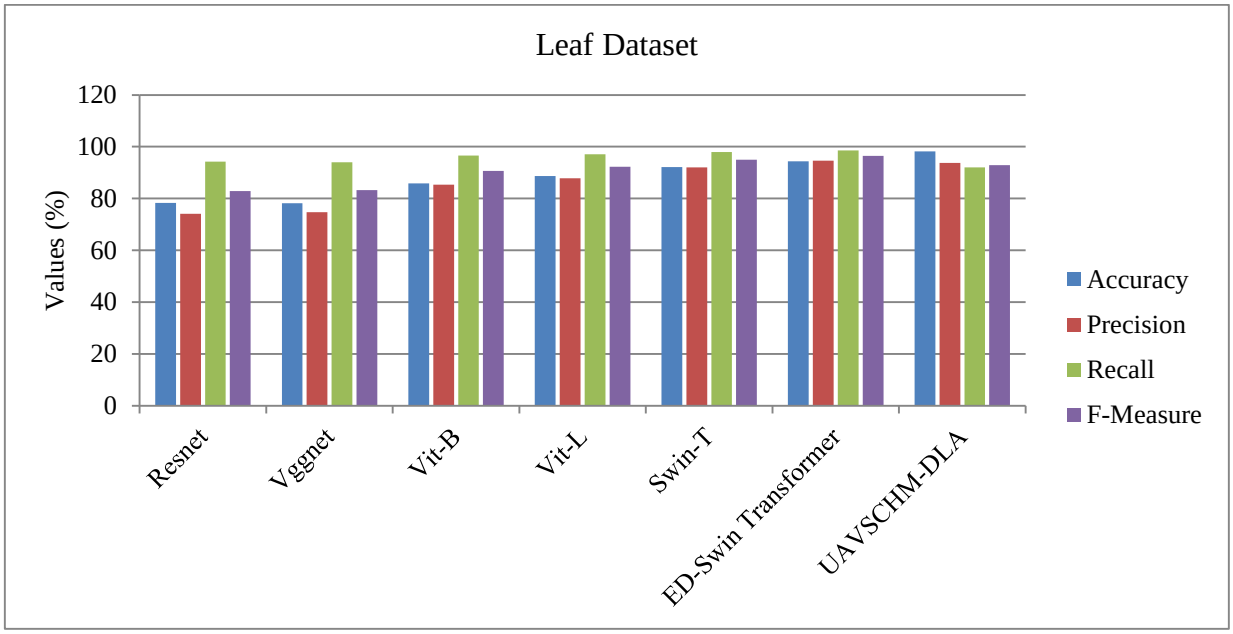

Fig. 8 Comparison of outcomes of UAVSCHM-DLA techniques with existing approaches on the leaf dataset

The UAV dataset includes 2845 aerial images organised into four representative classes, with their detailed distribution presented in Table 5.

Table 5. Details of the UAV dataset

UAV Dataset		
Class Name	**Labels**	**No. of Images**
Healthy	Class 1	281
Mosaic	Class 2	773
Rust	Class 3	1001
Caterpillar and Semi-Lopper Pest Attack	Class 4	790
Total		**2845**

Figure 9 presents the crop health monitoring results of the UAVSCHM-DLA methodology under a 70:30 TRAP/TESP ratio. Figure 9(a) reveals the confusion matrix of each class. Figure 9(b) shows the PR examination, indicating a better outcome across all class labels. Finally, Figure 9(c) shows the ROC curve, indicating that capable outcomes are associated with a higher ROC curve across diverse classes.

Table 6 and Figure 10 present the crop health monitoring results of the UAVSCHM-DLA approach on the UAV dataset with a 70:30 split. The outcome suggests that the UAVSCHM-DLA approach correctly recognised the instances. With 70%TRAP and 30%TESP, the UAVSCHM-DLA methodology presents an average $accur_y$, $preci_n$, $recal_l$, $F_{measure}$, and AUC_{score} of 96.89%, 93.51%, 92.81%, 93.14%,

and 95.31%, and 97.01%, 93.82%, 92.26%, 92.96%, and 95.09%, respectively.

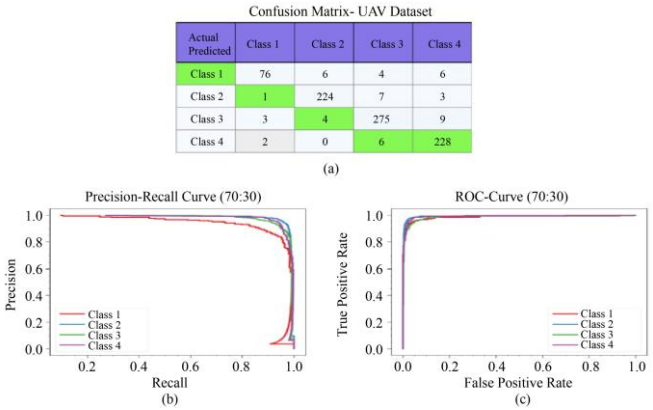

Fig. 9 (a) Confusion matrix, (b) PR, and (c) ROC curves under 70:30 on the UAV dataset.

Table 6. Crop health monitoring of the UAVSCHM-DLA approach under 70:30 on the UAV dataset

Class Labels	$Accur_y$	$Preci_n$	$Recal_l$	$F_{measure}$	AUC_{score}
TRAP (70%)					
Class 1	98.09	92.18	87.30	89.67	93.26
Class 2	97.64	95.38	95.91	95.64	97.10
Class 3	95.18	93.86	92.54	93.19	94.59
Class 4	96.63	92.64	95.49	94.04	96.28
Average	**96.89**	**93.51**	**92.81**	**93.14**	**95.31**
TESP (30%)					
Class 1	97.42	92.68	82.61	87.36	90.91
Class 2	97.54	95.73	95.32	95.52	96.85
Class 3	96.14	94.18	94.50	94.34	95.74
Class 4	96.96	92.68	96.61	94.61	96.85
Average	**97.01**	**93.82**	**92.26**	**92.96**	**95.09**

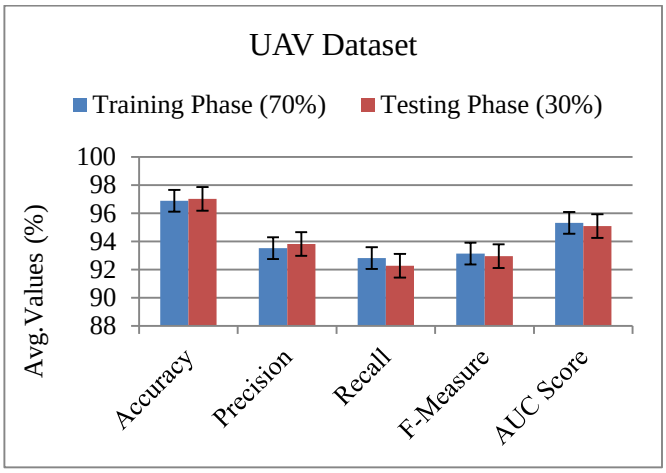

Fig. 10 Crop health monitoring of UAVSCHM-DLA approach under 70:30 on UAV dataset

Figure 11 reveals the training (TRAN) and validation (VALD) $accur_y$ of the UAVSCHM-DLA approach over 200 epochs. The slow convergence of the model indicates better generalization and no overfitting.

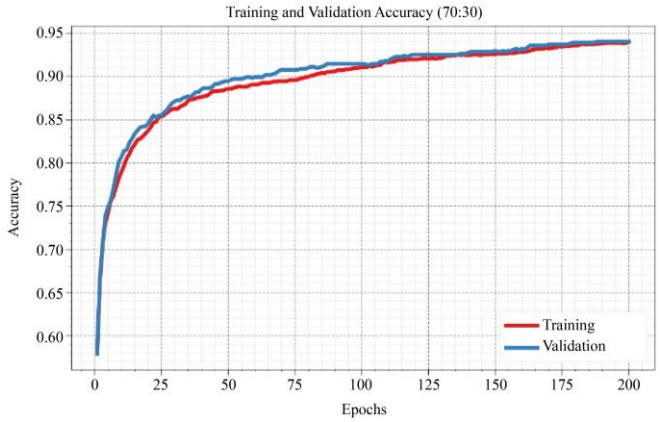

Fig. 11 $accur_y$ curve of UAVSCHM-DLA approach under 70:30 on UAV dataset

Figure 12 depicts the TRAN and VALD loss of the UAVSCHM-DLA technique over 200 epochs. The VALD remained lower than the TRAN loss throughout most epochs, emphasizing better generalisation and no overfitting.

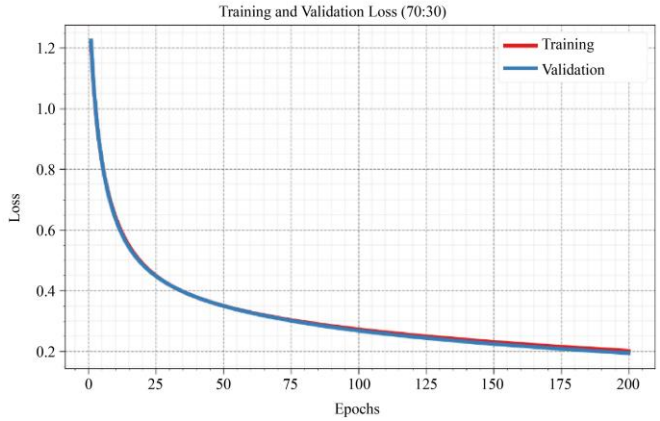

Fig. 12 Loss curve of UAVSCHM-DLA approach under 70:30 on UAV dataset

Table 7 and Figure 13 present the comparative results of the UAVSCHM-DLA methodology with existing approaches on the UAV dataset. The results illustrated that the UAVSCHM-DLA methodology achieved better results. Based upon $accur_y$, the UAVSCHM-DLA methodology has a superior $accur_y$ of 97.01%, while the Swin-S of 93.04%, VGG 16 + XGBoost of 85.03%, Full VGG 16 of 89.72%, VGG 16 + RF of 86.20%, VGG 16 + SVM of 96.71%, and VGG 16 + XGBoost of 91.60% got lesser outcome, correspondingly. Furthermore, based on $recal_l$, the UAVSCHM-DLA technique yields the best solution, with $a\ recal_l$ of 92.26%. In contrast, the Swin-S of 98.18%, VGG

16 + XGBoost of 87.00%, Full VGG 16 of 84.50%, VGG 16 + RF of 89.10%, VGG 16 + SVM of 96.10%, and VGG 16 + XGBoost of 92.20% provide minimal solutions, correspondingly. Ultimately, based on $F_{measure}$, the UAVSCHM-DLA technique accomplishes an enhanced solution with $an\ F_{measure}$ of 92.96%, although the Swin-S of 95.12%, VGG 16 + XGBoost of 86.20%, Full VGG 16 of 85.30%, VGG 16 + RF of 89.50%, VGG 16 + SVM of 96.00%, and VGG 16 + XGBoost of 91.20% present lower values, respectively.

Table 7. Comparison assessment of the UAVSCHM-DLA methodology with existing models on the UAV dataset

UAV Dataset				
Methods	$Accur_y$	$Preci_n$	$Recal_l$	$F_{measure}$
Swin-S	93.04	92.24	98.18	95.12
VGG 16 + XGBoost	86.20	86.00	87.00	86.20
Full VGG 16 model	85.03	85.10	84.50	85.30
VGG 16 + RF	89.72	89.60	89.10	89.50
VGG 16 + SVM	96.71	96.50	96.10	96.00
VGG 16 + XGBoost	91.60	91.10	92.20	91.20
UAVSCHM-DLA	97.01	93.82	92.26	92.96

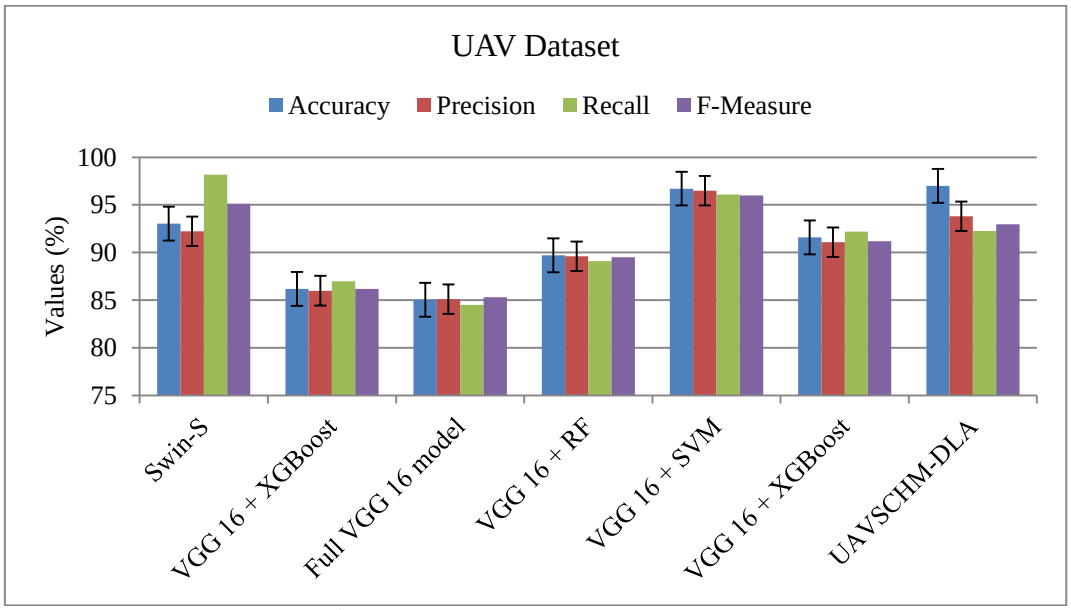

Fig. 13 Comparison assessment of the UAVSCHM-DLA methodology with existing models on the UAV dataset

From the detailed results and discussion, it is clear that the UAVSCHM-DLA technique achieves superior performance compared to other models.

5. Conclusion

In this manuscript, a novel UAVSCHM-DLA approach is presented to monitor and assess soybean crop health utilising integrated UAV and leaf images. Initially, the UAVSCHM-DLA approach performs image pre-processing using HE to enhance contrast and overall brightness, and BF is employed to remove noise from the imagery for further analysis. For effective feature representation, an IMViT is used to extract more discriminative, contextually rich representations from pre-processed images. Lastly, MNet-Attn is utilised for effective classification of soybean crop health conditions using the extracted rich feature representations. The comparison assessment of the UAVSCHM-DLA technique showed superior accuracies of 98.20% and 97.01% on the leaf and UAV datasets, respectively.

References

[1] Shanxin Zhang et al., "Monitoring of Soybean Maturity using UAV Remote Sensing and Deep Learning," *Agriculture*, vol. 13, no. 1, pp. 1-21, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Everton Castelão Tetila et al., "Detection and Classification of Soybean Pests using Deep Learning with UAV Images," *Computers and Electronics in Agriculture*, vol. 179, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Everton Castelão Tetila et al., "Automatic Recognition of Soybean Leaf Diseases using UAV Images and Deep Convolutional Neural Networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 5, pp. 903-907, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jayme Garcia Arnal Barbedo, "Deep Learning for Soybean Monitoring and Management," *Seeds*, vol. 2, no. 3, pp. 340-356, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Tej Bahadur Shahi et al., "Recent Advances in Crop Disease Detection using UAV and Deep Learning Techniques," *Remote Sensing*, vol. 15, no. 9, pp. 1-29, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Yu-Hyeon Park et al., "Detection of Soybean Insect Pest and a Forecasting Platform using Deep Learning with Unmanned Ground Vehicles," *Agronomy*, vol. 13, no. 2, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Pengting Ren et al., "Estimation of Soybean Yield by Combining Maturity Group Information and Unmanned Aerial Vehicle Multi-Sensor Data Using Machine Learning," *Remote Sensing*, vol. 15, no. 17, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Bo Zhang, and Dehao Zhao, "An Ensemble Learning Model for Detecting Soybean Seedling Emergence in UAV Imagery," *Sensors*, vol. 23, no. 15, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Abdelmalek Bouguettaya et al., "Deep Learning Techniques to Classify Agricultural Crops Through UAV Imagery: A Review," *Neural Computing and Applications*, vol. 34, pp. 9511-9536, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Maitiniyazi Maimaitijiang et al., "Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning," *Remote Sensing*, vol. 12, no. 9, pp. 1-23, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xun Yu et al., "FEL-YoloV8: A New Algorithm for Accurate Monitoring Soybean Seedling Emergence Rates and Growth Uniformity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 63, pp. 1-17, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Muhammad Aqeel et al., "Real-Time Crop Health Monitoring Using AI-Based Drone Surveillance and YOLOv12," *Pakistan Journal of Scientific Research*, vol. 5, no. 1, pp. 29-39, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Xiaoming Li et al., "HSDT-TabNet: A Dual-Path Deep Learning Model for Severity Grading of Soybean Frogeye Leaf Spot," *Agronomy*, vol. 15, no. 7, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Mashrur Kabir et al., "*Design and Implementation of a Crop Health Monitoring System*," Doctoral Thesis, Brac University, pp. 1-182, 2023. [Google Scholar] [Publisher Link]

[15] Sean Wallinger et al., "Toward a Cost-Effective Smart Crop Health Monitoring System," *2023 IEEE MetroCon*, Hurst, TX, USA, pp. 1-3, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Amelia Sarah Binti Abdul Rahman et al., "Multispectral Image Analysis for Crop Health Monitoring System," *2022 IEEE 5th International Symposium in Robotics and Manufacturing Automation (ROMA)*, Malacca, Malaysia, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Thangavel Murugan et al., "Research Advances in Maize Crop Disease Detection Using Machine Learning and Deep Learning Approaches," *Computers*, vol. 15, no. 2, pp. 1-53, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[18] Juntao Tong et al., "ToT-Net: A Generalized and Real-Time Crop Disease Detection Framework via Task-Level Meta-Learning and Lightweight Multi-Scale Transformer," *Smart Agricultural Technology*, vol. 12, pp. 1-17, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Muhammad Nouman Noor et al., "An Effective Approach for Recognition of Crop Diseases Using Advanced Image Processing and YOLO v8," *Food Science & Nutrition*, vol. 141, no. 2, pp. 1-26, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[20] V. Gopinath, and M. Sangeetha, "Spatial-Temporal Digital Twin for Crop Disease Prediction: A Hybrid ConvLSTM-Graph Neural Network Approach," *2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, pp. 570-575, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[21] Hardeep Kaur, Bhanu Priya, and Kuldeep Singh, "CNNx: Optimizing Smart CNN Models for Efficient Banana Disease Detection and Severity Estimation," *Concurrency and Computation: Practice and Experience*, vol. 38, no. 1, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[22] Vivek Parganiha, and Monika Verma, "An Efficient Disease Prediction in Smart Agriculture Using Advanced Deep Learning Methods for Improving Crop Productivity," *Journal of Phytopathology*, vol. 173, no. 5, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Anqi Kang et al., "A^2Former: An Airborne Hyperspectral Crop Classification Framework Based on a Fully Attention-Based Mechanism," *Remote Sensing*, vol. 18, no. 2, pp. 1-29, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[24] Zhen Du et al., "UniHSFormer X for Hyperspectral Crop Classification with Prototype-Routed Semantic Structuring," *Agriculture*, vol. 15, no. 13, pp. 1-32, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[25] Gourav Mondal, Rajesh Kumar Dhanaraj, and Md. Shohel Sayeed, "UAV-MCND: A Novel System for Multiclass Natural Disaster Classification Using FusionNet-4 and Water Wheel-Guided Walrus Optimization," *International Journal of Intelligent Systems*, vol. 2025, no. 1, pp. 1-25, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Qihao Chen et al., "IMViT: Adjacency Matrix-Based Lightweight Plain Vision Transformer," *IEEE Access*, vol. 13, pp. 18355-18545, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[27] Xiaobin Wei et al., "Improved MNet-Atten Electric Vehicle Charging Load Forecasting Based on Composite Decomposition and Evolutionary Predator–Prey and Strategy," *World Electric Vehicle Journal*, vol. 16, no. 10, pp. 1-23, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Sayali Shinde, and Vahida Attar, "An Indian UAV and Leaf Image Dataset for Integrated Crop Health Assessment of Soybean Crop," *Data in Brief*, vol. 60, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jing Zhang et al., "ED-Swin Transformer: A Cassava Disease Classification Model Integrated with UAV Images," *Sensors*, vol. 25, no. 8, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[30] Girma Tariku et al., "Advanced Image Pre-Processing and Integrated Modeling for UAV Plant Image Classification," *Drones*, vol. 8, no. 11, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]