

Advanced Crop Health Monitoring for Smart Agriculture Using Fusion of Transfer Learning With Attentive Convolutional Recurrent Neural Network on Remote Sensing Images

Amani K. Samha , Nada Alzaben, Mohammed Abaker , Faten Derouez, Wafi Bedewi, Mohammed Alahmadi, Asma A. Alhashmi , and Mutasim Al Sadig 

Abstract—Climate change and variability, highlighted by increasing temperatures, altered precipitation, and extreme weather, have significant effects on agriculture. A suitable crop assessment at an enhanced level needs observing extensive areas by a powerful method. Remote sensing (RS) skill provides this over nondestructive synoptic screening abilities. The technology of RS has revolutionized the farming sector by providing vital insights into resource allocation, environmental sustainability, and crop management. RS provides a cost-effective and noninvasive means of observing crops throughout their lifecycle. Early recognition allows for targeted interventions, decreasing crop damage and diminishing the requirement for chemical inputs. It also helps in

Received 30 July 2025; revised 12 September 2025; accepted 17 September 2025. Date of publication 29 September 2025; date of current version 23 October 2025. This work was supported in part by the Deanship of Research and Graduate Studies at King Khalid University through Large Research Project under Grant RGP2/569/46, in part by Princess Nourah bint Abdulrahman University Researchers Supporting Project under Grant PNURSP2025R733, in part by Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, in part by Ongoing Research Funding program, King Saud University, Riyadh, Saudi Arabia, under Grant ORF-2025- 550, in part by the Deanship of Scientific Research at Northern Border University, Arar, KSA under Project “NBU-FPEJ-2025- 2903-01, in part by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia, under Grant KFU253315, in part by the Deanship of Graduate Studies and Scientific Research at University of Bisha through the Fast-Track Research Support Program, and in part by the Deanship of Scientific Research at Majmaah University under Project ER-2025-2008. (*Corresponding author: Mutasim Al Sadig*)

Amani K. Samha is with the Department of Management Information Systems, College of Business Administration, King Saud University, Riyadh 28095, Saudi Arabia.

Nada Alzaben is with the Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia.

Mohammed Abaker is with the Department of Computer Science, Applied College at Mahayil, King Khalid University, Abha 62521, Saudi Arabia.

Faten Derouez is with the Department of Quantitative Method, College of Business Administration, King Faisal University, Al Ahsa 31982, Saudi Arabia.

Wafi Bedewi is with the Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia.

Mohammed Alahmadi is with the Department of Cybersecurity, College of Computer Science and Engineering, University of Jeddah, Jeddah 21493, Saudi Arabia.

Asma A. Alhashmi is with the Department of Computer Science, College of Science, Northern Border University, Arar 73213, Saudi Arabia.

Mutasim Al Sadig is with the Department of Computer Science, College of Science, Majmaah University, Al Majmaah 11952, Saudi Arabia (e-mail: mm.elkir@mu.edu.sa).

Digital Object Identifier 10.1109/JSTARS.2025.3615906

monitoring an environment. In the era of digital data propagation, agriculture is poised for a transformative innovation driven by deep learning, utilizing RS imagery. In this manuscript, a robust crop health assessment using an improved fusion neural network with remote sensing imagery (RCHA-IFNNRSI) approach is proposed. The aim is to develop an integrated crop health assessment framework using RS and leaf image analysis for enhancing precision and decision-making in smart agriculture. In the initial image preprocessing stage, the bilateral filter based noise elimination and contrast enhancement is performed to enhance the raw images by removing noise. For feature extraction, the RCHA-IFNNRSI approach integrates swin transformer (ST) and ResNet50 models to create a fused backbone that effectively captures both local and global features for crop health assessment. Finally, the attentive convolutional recurrent neural network technique is employed to detect and classify crops into diverse health status classes automatically. Extensive experimentation of the RCHA-IFNNRSI technique on the MH-SoyaHealthVision dataset illustrated superior performance, achieving a high accuracy of 98.97% compared to existing models.

Index Terms—Crop health assessment, deep learning (DL), remote sensing (RS) imagery, smart agriculture, using improved fusion neural network.

I. INTRODUCTION

AGRICULTURE, an engine of economic growth for many countries, offers the most crucial needs of people: food and fibre. Technological transformations over the past several years, such as the Green Revolution, have changed the face of agriculture [1]. The enhanced crop varieties, synthetic pesticides, fertilizers, and irrigation improve food security and crop productivity, particularly in emerging countries. Therefore, even though doubling the population and increasing food demand, global agriculture can satisfy the needs with only 30% growth in the area under cultivation [2]. The need for agricultural and food products is expected to rise by more than 70%. Considering the constrained availability of arable land, a significant part of this increased need will be met by agricultural growth, i.e., improved utilization of fertilizer, water, pesticides, and other inputs [3]. There is a variety of features to monitor crop health, ranging from plant vigour, soil humidity, and accessibility, to stress provided by abiotic factors, such as temperature, humidity, and rainfall, and also biotic factors, such as disease and pest. Any delay from

the standard development parameter impacts crop growth and eventually diminishes productivity [4]. For this reason, it is vital to analyze crop health for the whole development cycle.

Suitable and appropriate crop evaluation at a high level necessitates observing broad regions with an effective system. RS and leaf image analysis are transforming crop health assessment in smart agriculture by offering early, precise, and detailed insights into plant health, allowing prompt actions and enhanced resource management [5]. RS, utilizing data from drones, satellites, and other platforms, provides a broader perspective of crop health. In contrast, leaf image analysis, frequently integrating machine learning (ML), gives comprehensive data regarding specific plants and possible issues [6]. This integrated method enables a more general and effective approach for managing crops, aiding in better sustainability and productivity. Conventionally, crop health assessment is based on visual examination, which is a longer, biased, and restricted process to noticeable symptoms [7]. Conversely, RS can find minor changes in plant health indicators, namely chlorophyll content and water stress, before they become evident to the naked eye. This early identification enables targeted interventions, such as using specific pesticides or fertilizers only where needed, instead of blanket treatments, minimizing costs and reducing environmental implications [8].

Leaf image analysis, which often employs ML and deep learning (DL) approaches, can additionally enhance this assessment by detecting specific diseases, nutrient deficiencies, or other problems at the leaf level [9]. The integration of RS and leaf image analysis in smart agriculture provides numerous benefits in crop health assessment through DL techniques. It enables real-time observation of crop health in broader areas, allowing farmers to detect and solve possible issues beforehand. These technologies also enable precision agriculture, in which inputs, including fertilizer, water, and pesticides, are applied only where necessary, enhancing resource use and decreasing waste [10]. By merging these technologies with other smart farming applications, namely data analytics and crop modeling, farmers can make optimal decisions, enhance yields, and improve the sustainability of their operations. Furthermore, the combination of RS and ML is transforming crop health assessment in smart agriculture by offering real-time monitoring, nondestructive, and early detection of issues [11]. This enables precision farming practices, boosting resource utilization and increasing yields.

In this manuscript, a robust crop health assessment using an improved fusion neural network with remote sensing imagery (RCHA-IFNNRSI) approach is proposed. The aim is to develop an integrated crop health assessment framework using RS and leaf image analysis for enhancing precision and decision-making in smart agriculture. In the initial image preprocessing stage, the bilateral filter (BF)-based noise elimination and contrast enhancement is performed to enhance the raw images by removing noise. For feature extraction, the RCHA-IFNNRSI approach integrates swin transformer (ST) and ResNet50 models to create a fused backbone that effectively captures both local and global features for crop health assessment. Finally, the attentive convolutional recurrent neural network (ACRNN) technique is employed to detect and classify crops into diverse health status classes automatically. The experimentation of the RCHA-IFNNRSI technique is performed on the MH-SoyaHealthVision

dataset. The significant contribution of the RCHA-IFNNRSI technique is listed below.

- 1) The BF model is used for preprocessing RS images to remove noise and improve contrast, resulting in more precise and more reliable data for analysis. This also assists feature extraction by providing high-quality images and strengthens the overall performance of the ACRNN in accurately evaluating crop health.
- 2) The fused backbone incorporating ST and ResNet50 is developed for effectively capturing both local details and global contextual data from RS images. This integration utilizes the merits of both architectures, improving feature representation for intrinsic crop health patterns. By improving feature extraction, it enables more robust and accurate analysis of crop conditions across diverse environments.
- 3) The ACRNN technique is employed to automatically classify crop health by efficiently capturing spatial and temporal dependencies. This also improves the model's ability to focus on crucial features while considering sequential data, resulting in more precise and reliable health assessments. It ultimately enhances the automation and accuracy of crop monitoring in smart agriculture.
- 4) A novel hybrid DL process integrating advanced image preprocessing techniques with a fused feature extraction backbone and attention-based temporal modelling is proposed. By incorporating these components, the approach efficiently captures both spatial and temporal aspects of crop health from RS images. This enables more precise and comprehensive monitoring, addressing limitations of prior methods in smart agriculture applications.

II. RELATED WORKS ON ROBUST CROP HEALTH ASSESSMENT IN SMART AGRICULTURE

Yang et al. [12] presented a new smart analysis scheme. It utilizes RS data from drones and a convolutional neural network (CNN) technique to attain effective crop classification and precise detection of diseases and pests. Moreover, this scheme applies multiscale attention convolutional networks to enhance crop detection, advances cyclic consistent adversarial networks for RS image translation, and improves the lightweight MobileNetV2 for pest and disease identification. This provides technological assistance for the transition to smart and sustainable agriculture. Ibrahim et al. [13] proposed an AI-IoT smart agriculture pivot as a good candidate for identifying and managing plant disease without relying on robotic or drone technology. Therefore, it provides a novel IoT framework and a hardware pilot, leveraging the current central pivot to develop DL methods for detecting plant diseases in various crops and controlling their actuators for disease treatment. In [14], AAVs equipped with a comprehensive soil and crop sensor package are utilized to examine the soil quality, crop health, and moisture level of an area. The AAVs can gather accurate data, thereby minimizing the time required to investigate plots and design a field management plan. De Ocampo and Montalbo [15] proposed a multivision monitoring approach that can concurrently recognize farm activities and assess crop condition. This comprises computer

vision methods that convert aerial videos into a sequence of images to identify significant environmental characteristics. This approach has dual key modules: the farmer activity recognition (FAR) module and the crop image analysis (CIA) module. FAR presents a new feature extraction technique that extracts motion in several maps, allowing discrete feature sets for all activities. In the meantime, the authors estimated the leaf chlorophyll levels, a key indicator of crop health. Munaganuri and Rao [16] developed an ML technique by integrating grey wolf optimised nonlocal means (GWO-NLM) for image denoising, the dual elephant herding optimisation (DEHO) method for feature selection, and the deep dyna Q graph convolutional network (DDQGCN) approach for classification.

Swarnkar et al. [17] examined the incorporation of artificial intelligence (AI) in smart agriculture to improve crop health observation and nutrient management. Employing AI to advance the sustainability and efficacy of farming practices. This study investigates the use of AI for early recognition of crop pests and diseases, and the progress of predictive models for accurate nutrient management. Imaging and RS techniques play a vital part in monitoring, enabling well-informed decisions. Ilyas et al. [18] utilized current precision agriculture technologies to enhance remote crop yield prediction and crop type classification, employing a fuzzy hybrid ensembled classification and prediction technique based on RS information. This investigation recognizes the optimum weights of the robust candidate classifiers for the ensembled classification technique, applying the bagging approach. Traditional yield prediction and crop classification approaches are ineffective under adverse conditions. In [19], an approach to optimizing ML's hyperparameters is suggested. The process of developing the optimum construction through the optimal configuration of hyperparameters is referred to as hyperparameter tuning. It is widely acknowledged that tuning hyperparameters is crucial for creating a robust ML approach. The need to enhance the effectiveness of a learning model is vital for effective utilization of an ML model. Salam et al. [20] utilized environmental factors such as relative humidity, soil moisture, temperature, light intensity, nitrogen, phosphorus, potassium (NPK), and pH to localize and cluster various autonomous aerial vehicles (AAVs) for precise detection of affected tomato crop areas. The model also utilizes the honey bee optimization (HBO) technique for improving detection. Fan and Chung [21] proposed a model by utilizing high-resolution AAV imagery. It also combines the visible atmospherically resistant index (VARI) with DL models, including single shot multibox detector (SSD), You Only Look Once version 3 (YOLOv3), and faster region-based CNN (faster RCNN) approach to enhance detection accuracy in visually complex environments.

Lacerda et al. [22] presented a technique by utilizing agriculture intelligence generative adversarial networks (AgI-GAN) model. The study also employs the Agroview platform. MirhoseiniNejad et al. [23] developed a robust soybean yield prediction model by integrating convolutional long short-term memory (ConvLSTM), three-dimensional CNN (3-D-CNN), and vision transformer (ViT) models. Zhang et al. [24] investigated the application of CNN, ViT, and CNN-transformer hybrid models for intelligent pest and disease detection. Farmonov et al.

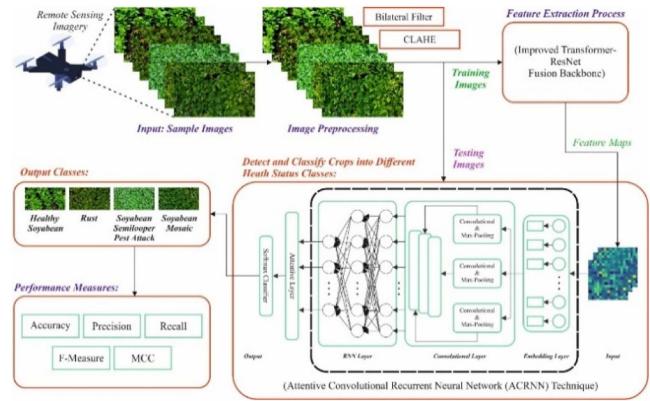


Fig. 1. Overall process of the RCHA-IFNNRSI model.

[25] proposed a DL methodology, namely HypsLiDNet, by incorporating hyperspectral imaging (HSI) and light detection and ranging (LiDAR) data by utilizing morphological feature extraction and an attention mechanism (AM) model. Kumar et al. [26] proposed a MobileNet-gated recurrent unit (MobileNet-GRU) fusion model. The technique comprises MobileNet-based spatial feature extraction, a lightweight CNN, and temporal pattern learning via the gated recurrent unit (GRU) method to improve real-time agricultural monitoring. Mahmoodi et al. [27] enhanced hyperspectral image (HSI) classification by utilizing principal component analysis (PCA) for dimensionality reduction, integrated with spectral and spatial attention modules to improve feature extraction. A hybrid neural network incorporating 3-D CNNs and two-dimensional CNNs (2-D CNNs) with skip connections is used to effectively capture both spectral and spatial information while mitigating model complexity. Hajare and Rajawat [28] presented a method by integrating deep belief network (DBN), recurrent neural network (RNN), and CNN within an ensemble framework. The model is optimized by utilizing a novel Swarm Intelligence-based Self-Improved Dwarf Mongoose Optimisation Algorithm (SIDMO) technique to enhance plant disease detection accuracy. Wang et al. [29] proposed a novel semantic segmentation model called Unified U-Net (UniU-Net) technique by utilizing auxiliary encoder, unified attention fusion module (UAFM), spatial AM, and structural reparameterization to improve segmentation accuracy. Summary of present models for crop health assessment in Table I.

III. PROPOSED SYSTEM ARCHITECTURE

This article develops an RCHA-IFNNRSI system. The primary purpose of this study is to create a combined crop health assessment framework utilizing RS and leaf image analysis to improve precision and decision-making in smart agriculture. To obtain this, the presented RCHA-IFNNRSI system has dual image preprocessing, dual feature extraction, and classification models. Fig. 1 indicates the overall process of the RCHA-IFNNRSI model.

TABLE I
COMPARISON STUDY OF DIFFERENT ADVANCED MODELS

Authors	Goals	Techniques	Results
Yang et al. [12]	Presents a novel intellectual analysis system. The system utilizes RS data from drones to accomplish effective crop classification and precise identification of diseases and pests.	CNN, Cycle GAN, and VGG16	Average F1 and intersection ratio of 94.67% and 89.14%
Ibrahim et al. [13]	Projects that involve plant disease identification and treatment depend on smart Agriculture for plant-based crop production in the field of agriculture.	DL and ResNet50	Accuracy of 99.8%
Young et al. [14]	UAVs equipped with whole crop and soil sensor packages are employed to assess the crop health, quality of soil, and moisture level of the field. It is capable of gathering accurate data and reducing the time required to examine the plots, thereby advancing the field management plan.	—	—
De Ocampo and Montalbo [15]	Presents an innovative method to estimate crop health and monitor farmer activities in a single framework. It focuses on extracting the requirements for many pipelines to carry out diverse monitoring tasks.	—	mAP of 87.02% and 94.48%
Munaganuri and Rao [16]	To project a new ML methodology that precisely approximates the crop's water requirements and enhances irrigation scheduling using RS sample data.	GWO-NLM, DEHO, DDQGCN	Accuracy of 8.3% and Precision of 8.5%
Swarnkar et al. [17]	To develop the employment of AI for prior identification of crop disorders and pests, along with the enhancement of predictive techniques for accurate nutrient management. RS and imaging technologies perform a vital role in real-time monitoring, permitting informed decision-making.	AI	—
Ilyas et al. [18]	To leverage modern precision agriculture tools for improved remote crop yield assessment, it is essential to recognize the optimal weight of the strongest candidate classifiers in the ensembled classification methodology, thereby accepting the bagging approach.	Ensemble Models	Average of 13% and 24%
Venkatakotireddy et al. [19]	Presents a model to optimize the hyperparameters of conventional ML. The process of intending the optimum perfect construction through the optimum configuration of hyperparameters.	ML	—
Salam et al. [20]	To localize and cluster various UAVs for the accurate detection of affected tomato crop areas.	Weight-based UAV Localisation, HBO, Cluster Formation	Communication Overhead, Packet Delivery Ratio, End-To-End Delay, Energy Consumption
Fan and Chung [21]	To enhance automated banana detection in intercropped fields using UAV imagery and VARI-enhanced DL.	SSD, YOLOv3, Faster RCNN	Precision, Recall, intersection over union (IoU), Training Time
Lacerda et al. [22]	To improve image resolution using AgI-GAN to match UAV-level quality for precision agriculture applications.	AgI-GAN, Integration with Agroview	Detection error, Cost per acre
MirhoseiniNejad et al. [23]	To develop an integrated model using multispectral RS data.	ConvLSTM, 3D-CNN, ViT	Root Mean Square Error, Correlation Coefficient
Zhang et al. [24]	To review and analyze the application of conventional ML and DL models for pest and disease detection.	CNN, ViT, and CNN-Transformer	Accuracy, generalization ability, and model efficiency
Farmonov et al. [25]	To develop a DL technique by integrating HSI and LiDAR data for improved crop classification accuracy.	HypsLiDNet, HSI, LiDAR, AM	Classification Accuracy, Computational Efficiency
Kumar et al. [26]	To enhance sustainable agriculture and food security on RS data.	MobileNet-GRU	Accuracy, Mean Absolute Error (MAE)
Mahmoodi et al. [27]	To enhance HSI classification.	PCA, Entropy-based informative module, Spectral attention module, Depthwise spatial attention block, Hybrid 3-D and 2-D CNN with skip connections	Classification Accuracy, Feature Extraction Quality
Hajare and Rajawat [28]	To develop an automated deep ensemble model for the accurate classification of black gram crop diseases.	DBN, RNN, CNN, SIDMO	Accuracy, Precision, Recall
Wang et al. [29]	The goal is to develop a high-precision segmentation model in complex remote sensing (RS) imagery.	UniU-Net, UAFM	F1-score, IoU

A. Image Preprocessing Techniques

In the beginning, the image preprocessing step includes BF-based noise removal and CLAHE-based contrast enhancement to improve the raw images by eliminating noise. This is chosen for its efficiency in mitigating noise while also preserving crucial edge and fine image details. Unlike conventional smoothing filters that may blur crucial features, BF maintains spatial integrity, ensuring that relevant textures and boundaries remain intact. Furthermore, CLAHE improves local contrast adaptively, enhancing visibility in regions with varying illumination without over-amplifying noise. This incorporation also outperforms standard global contrast enhancement models. It provides a balanced approach that prepares high-quality inputs for DL models, ultimately improving detection and classification accuracy compared to other preprocessing techniques.

1) *BF-Based Noise Removal*: The BF is an extensively applied method for noise reduction in image processing [30].

It successfully smoothes out noise, maintains significant edge details, making it highly relevant to improving image quality before analysis

$$B(I)_p = \frac{1}{W_p} \sum_{q \in \Omega} I_q \cdot g_{\sigma_s}(\|p - q\|) \cdot g_{\sigma_r}(|I_p - I_q|). \quad (1)$$

- The BF is used to process the input image I , yielding a filtered image $B(I)$. The output value at each pixel p is determined by incorporating either spatial or intensity information.
- Particularly, $(I)_p$ and $(I)_q$ characterize the intensity values at pixels p and q in the new image.
- The filter works over the neighborhood Ω around every pixel p .
- A spatial Gaussian function $g_{\sigma_s}(\|p - q\|)$ explains the geometric closeness among p and q , whereas a range Gaussian function $g_{\sigma_r}(|I_p - I_q|)$ assesses their intensity similarity. This dual consideration guarantees that edges are preserved while smoothing areas with the same intensity.
- The normalization W_p guarantees proper scaling of the filtered outcome. Once applied to images, this model successfully lowers noise and preserves essential features.

2) *CLAHE-Based Contrast Enhancement*: CLAHE is a model for image development that improves contrast without increasing noise excessively [31], whereas normal HE uses the uniform change to the complete image, CLAHE reallocates intensity values locally by dealing with smaller sections named tiles. These assurances ensure the development of small details without going overboard with darkness or brightness. CLAHE also implements a contrast-limiting parameter to avoid over-saturation in regions with significant intensity variations. It uses a transformation function according to the cumulative distribution function for every tile's histogram

$$I_{\text{CLAHE}}(x, y) = I_{\min} + \frac{(I(x, y) - I_{\min})}{I_{\max} - I_{\min}} \times I'_{\max} - I'_{\min} \quad (2)$$

whereas $I_{\text{CLAHE}}(x, y)$ characterizes the output pixel intensity after using CLAHE, $I(x, y)$ denotes the original pixel intensity in the provided tile, I_{\max} and I_{\min} represent maximum and

minimum intensity values inside the local tile, and I'_{\min} and I'_{\max} describe the chosen intensity range after using CLAHE.

B. Feature Extraction Using Dual Backbone Methods

For the feature extraction process, the model utilizes two generally used backbones, ST and ResNet-50 [32]. The ST model is chosen for its excellence in capturing global context and long-range dependencies through its self-AM, which is significant for comprehending intrinsic spatial relationships in RS images. In addition, ResNet50 extracts local features and fine details, ensuring robust representation of texture and shape. This integrated model provides a better understanding of both local and global image features, and also enhances accuracy and robustness in diverse agricultural environments, making it an ideal choice for precise crop health assessment.

The significant development of the network presented in this work concentrates on the encoding portion. Higher-resolution images naturally have features of dissimilar scales, comprising complete outlines and local particulars, which are challenging to capture simultaneously, utilizing a single backbone. This work uses two generally applied backbones, ResNet50 and ST, to cooperatively construct a combined backbone with the strongest feature extraction abilities.

1) Basic Backbone:

a) *ResNet*: ResNet50 is a deep CNN derived from the ResNet structure that is one of the most commonly utilized backbone methods. The residual blocks in ResNet, which apply skip connections, permit layers that degrade performance to be bypassed over these links, successfully avoiding the degradation problem that occurs with improving networking depth. It includes four phases, each consisting of some residual blocks. Initially, in every phase, the feature mapping dimension is halved, whereas the channel counts are doubled. Every residual block comprises dual 3x3 convolutional layers for feature extraction. After every convolutional layer, the rectified linear unit (ReLU) activation function and batch normalisation (BN) are applied [33]. Fig. 2 signifies the structure of the ResNet-50 model.

b) *Swin transformer*: ST is a DL method derived from the Transformer structure, which redesigns the transformer architecture utilizing the model equivalent to CNNs. Afterwards, the image input undergoes patch partitioning, accompanied by linear embedding, and then enters the ST Block. Comparable with ResNet50, the ST includes four phases, each of which lowers the feature mapping resolution while improving the channel counts. Patches slowly merge and attain a pyramid architecture similar to CNNs. The ST block contains window-based multihead self-attention and shifted window MSA. ST presents hierarchic feature mapping and windowed attention transformation, addressing two key tasks associated with using transformers. These tasks consist of processing significant visual entity variations through dissimilar scenes, and handling computational efficiency related to global

2) *Fusing Strategies*: This work generated two significant developments in fusion approaches: i) presenting a two-information-interactive fusion structure, and ii) substituting conventional element-wise summation with the FS attention block for the fusion approach. Conventional fusion approaches

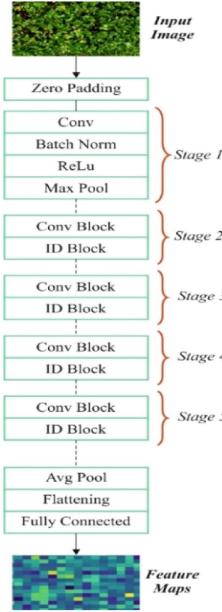


Fig. 2. Structure of the ResNet-50 model.

frequently accept a lead-assisting structure, where one backbone is measured as the leading backbone and the other as the assisting backbone. The assisting backbone works independently, whereas the leading backbone combines the features output by the consistent layers of the assisting backbone with its own features before proceeding to the next phase. This model failed to allow information exchange among the dual branches, and the assisting backbone can get information from the lead backbone

$$M_l = F(M_{sl} M_{rl}) \quad (3)$$

whereas M_l signifies the fused feature maps at the 1st phase, while M_{sl} and M_{rl} signify the feature mapping output by ST and ResNet50 at the 1st phase, correspondingly. F characterizes the fusion function

$$M_{sl+1} = \text{Swinlayer}_l(M_l). \quad (4)$$

Here, M_{sl+1} denote the feature mapping output by the ST at the $(1+1)$ th phase, and Swinlayer_l characterizes the ST layer at the 1st phase

$$M_{rl+1} = \text{Reslayer}_l(M_l). \quad (5)$$

Now, M_{rl+1} refers to the feature mapping output by the ResNet50 at the $(1+1)$ th phase, and Reslayer_l characterizes the ResNet50 layer at the 1st phase.

Consider that transformer outshines at removing global information, whereas the ResNet is improved at taking local information; the significance of the features removed by the two methods can vary across different scales. Directly addressing together for fusion failed to fine-tune the feature maps' weights adaptively, thus avoiding the attainment of successful features at dissimilar phases. Then, the FS block is used to fuse the feature mapping of ST and ResNet50 at every step. FS block is the attention module established according to the attention gate mechanism, which includes dual elements: "Fuse" and "Select." During the "Fuse"

step, two equally sized feature maps are initially fused and then reduced dimensionally. In the "Select" step, the softmax function is used to selectively weight the features of consistent channels, which are then multiplied back with the new feature mapping.

Fuse: Features from the similar phase of ResNet50 and ST are incorporated by element-wise summary to

$$U = U_r + U_s \quad (6)$$

whereas U denotes the incorporated feature, U_r and U_s signify the feature mapping output by ResNet50 and ST, respectively. This combined feature U is then passed over global average pooling F_{gp} , leading to the feature S , whereas $S \in R^c$, and the c th channel feature S_c of S is computed as shown

$$S_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j). \quad (7)$$

Then, s_c passes over the fully connected (FC) layer F_{fc} for feature selection, giving a feature vector Z of dimension d , as defined by the succeeding equation

$$Z = F_{fc}(S) = \delta(B(Ws)) \quad (8)$$

whereas $Z \in R^{d \times 1}$, δ signifies the activation function of RELU, B indicates BN, and $s \in R^{d \times 1}$. The dimensionality d is measured by the coefficient r , using the following equation:

$$d = \max\left(\frac{C}{r}, L\right). \quad (9)$$

Here, L signifies the minimal value of d ; L is fixed to a generally applied value of 32.

Select: Dual FC layers are applied for mapping z back to the new C -dimensional area, giving two feature vectors. Next, the channel sizes of the two feature vectors are handled utilizing the Softmax function to finish the channel FS for the dissimilar receptive fields

$$a_c = \frac{e^{A_{cz}}}{e^{A_{cz}} + e^{B_{cz}}} \quad (10)$$

$$b_c = \frac{e^{B_{cz}}}{e^{A_{cz}} + e^{B_{cz}}} \quad (11)$$

whereas $A_t B \in R^{c \times d}$, and a and b signify the soft attention weights for features U_r and U_s , correspondingly. $A_c \in R^{1 \times d}$ denotes the c th matrix row A , and a_c represents the attention weight set to the c th channel.

At last, the attention weights are multiplied by the new features to get the c th channel feature V_c , as demonstrated

$$V_c = a_c \cdot U_{rc} + b_c \cdot U_{sc} \quad (12)$$

$$a_c + b_c = 1. \quad (13)$$

The last comprehensive fusion feature maps representation is $V = [V_1, V_2, \dots, V_c], V_c \in R^{H \times W}$.

In comparison with these three methods, according to three other backbone network fusion strategies. The detailed information is as demonstrated.

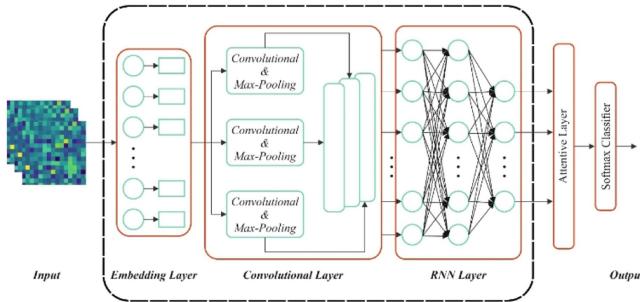


Fig. 3. Architecture of the ACRNN model.

- 1) ST helps ResNet50 lead strategy (SARL). This strategy fuses the output features from the similar phase of either backbone, with ST as the assisting backbone and *ResNet50* as the leading backbone.
- 2) ResNet50 helps ST lead strategy. This strategy is consistent in structure with the SARL, but exchanges the roles of the two backbones.
- 3) Two information-interactive composite strategy (DIC). This strategy assigns equivalent roles to both backbones.

C. ACRNN-Based Classification Process

At last, the ACRNN model is applied to automatically identify and classify crops into dissimilar health status classes [34]. This method is chosen for its capability in capturing both spatial and temporal dependencies in crop health data. The spatial features are efficiently extracted by the convolutional layers, while the recurrent layers model sequential patterns and changes over time, which are significant for monitoring crop growth and health progression. The interpretability and classification accuracy are also improved by the AM model, thus enabling precise and automated classification of diverse crop health conditions, making it highly effective for real-world agricultural monitoring. After choosing the feature condition, the removed features were transferred to the ACRNN method that incorporated convolutional, recurrent, and AM to process the composite time-frequency architecture. Fig. 3 indicates the architecture of the ACRNN model.

Convolutional layers: A sequence of convolutional layers removed local temporal and spectral designs [35]. For all layers l , the process of convolution was described as

$$X_{i,j}^{(l)} = \sigma \left(\sum_{p,q} W_{p,q}^{(l)} X_{i-p,j-q}^{(l-1)} + b^{(l)} \right) \quad (14)$$

(modified from the typical convolution process) whereas $b^{(l)}$ and $W^{(l)}$ signified biases and the learnable filters, and σ denotes the nonlinear activation, specifically the ReLU.

Recurrent layers: GRU is applied for capturing longer-term temporal dependences. A GRU's hidden layer (HL) upgrades are presented below

$$\begin{aligned} z_t &= \sigma (W_z h_{t-1}, x_t] + b_z \\ r_t &= \sigma (W_r h_{t-1}, x_t] + b_r \end{aligned} \quad (15)$$

TABLE II
DETAILS OF THE DATASET

Type of Dataset	Disease and Pest attack	Number of Images
UAV-based	“Healthy”	280
	“Mosaic”	772
	“Rust”	1000
	“Caterpillar and Semi lopper Pest Attack (CSLPA)”	790
	Total Images	2842

$$\begin{aligned} \tilde{h}_t &= \tanh (W_h r_t \odot h_{t-1}, x_t] + b_h \\ h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \end{aligned} \quad (16)$$

(expressed according to the GRU mechanism) whereas z_t and r_t represented the update and reset gates; however, h_t summarised the data required for classification.

Attention Mechanism: An attention layer is particularly concentrated on the pertinent time frames. For HL $\{h_1, \dots, h_T\}$, the attention score α_t were calculated as

$$e_t = v^T \tanh (W_a h_t + b_a), \alpha_t = \frac{\exp(e_t)}{\sum_{m=1}^T \exp(e_m)} \quad (17)$$

whereas W_a , b_a , and v are learnable parameters. The last contextual vector c aggregates HLs depending on α_t

$$c = \sum_{t=1}^T \alpha_t h_t \quad (18)$$

(according to additive attention).

This vector c was then transferred to the FC layer to provide the last activation vector z . Finally, a softmax function was applied to z to map it into the likelihood distribution \hat{y} , specifically through class C

$$\hat{y}_k = \frac{\exp(zk)}{\sum_{j=1}^C \exp(Zj)}. \quad (19)$$

IV. VALIDATION AND RESULTS

A. Dataset Description

In this section, the experimental validation outcomes of the RCHA-IFNNRSI method are under the MH-SoyaHealthVision dataset [36], [37]. It is a detailed dataset used for integrated crop health assessment in soybean farming. It integrates AAV-captured images and ground-level leaf images, allowing a holistic approach for detecting diseases and pest attacks. In addition to this, AAV datasets offer large-scale aerial views of a soybean field, identifying patterns of mosaic virus, rust, and pest infestations. This dataset focuses on enhancing crop health observation, enabling early intervention tactics and improving productivity in soybean farming. It contains 2842 images of four diseases and pest attacks, as shown in Table II. Fig. 4 illustrates the sample images of four classes. The RCHA-IFNNRSI technique is simulated using Python 3.6.5 on a PC with an i5-8600k, 250GB SSD, GeForce 1050Ti 4 GB, 16 GB RAM, and 1 TB HDD.

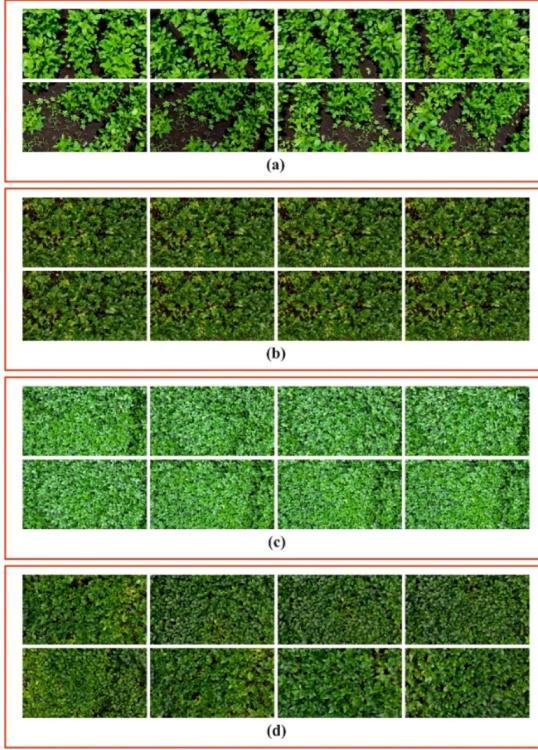


Fig. 4. Sample images of four classes. (a) Healthy soyabean. (b) Rust. (c) Soyabean semilooper pest attack. (d) Soyabean mosaic.

Parameters include a learning rate of 0.01, ReLU activation, 50 epochs, 0.5 dropout, and a batch size of 5.

B. Result Analysis

Fig. 5 shows the classifier outcomes of the RCHA-IFNNRSI approach at 80:20 and 70:30. Fig. 5(a), (d), (e)–(h) display the confusion matrices with accurate detection and classification of each class. Fig. 5(b) and (f) exhibit the PR inspection, denoting maximal performance in each class. At last, Fig. 5(c) and (g) portray the ROC investigation, exhibiting efficacious outcomes with superior ROC values for separate classes.

Table III and Fig. 6 exemplify the classifier outcome of the RCHA-IFNNRSI approach on 80:20. Based on 80% of the training phase (TRPHE), the RCHA-IFNNRSI model obtains an acc_{y} , prec_{n} , reca_{l} , F_{Measure} , and MCC of 98.97%, 97.96%, 97.01%, 97.47%, and 96.77%, respectively. At 20% of the testing phase (TSPHE), the RCHA-IFNNRSI model obtains an acc_{y} , prec_{n} , reca_{l} , F_{Measure} , and MCC of 98.33%, 96.47%, 96.42%, 96.43%, and 95.29%, respectively.

Fig. 7 exemplifies the training (TRAIN) acc_{y} and validation (VALID) acc_{y} of the RCHA-IFNNRSI method on 80:20 over 25 epochs. At first, both TRAIN and VALID acc_{y} rise rapidly, denoting efficient pattern learning from the data. Around the epoch, the VALID acc_{y} minimally exceeds the training accuracy, indicating good generalization without overfitting. As training advances, it reflects maximum performance and a minimum performance gap between TRAIN and VALID. The close alignment of both curves in training indicates that the

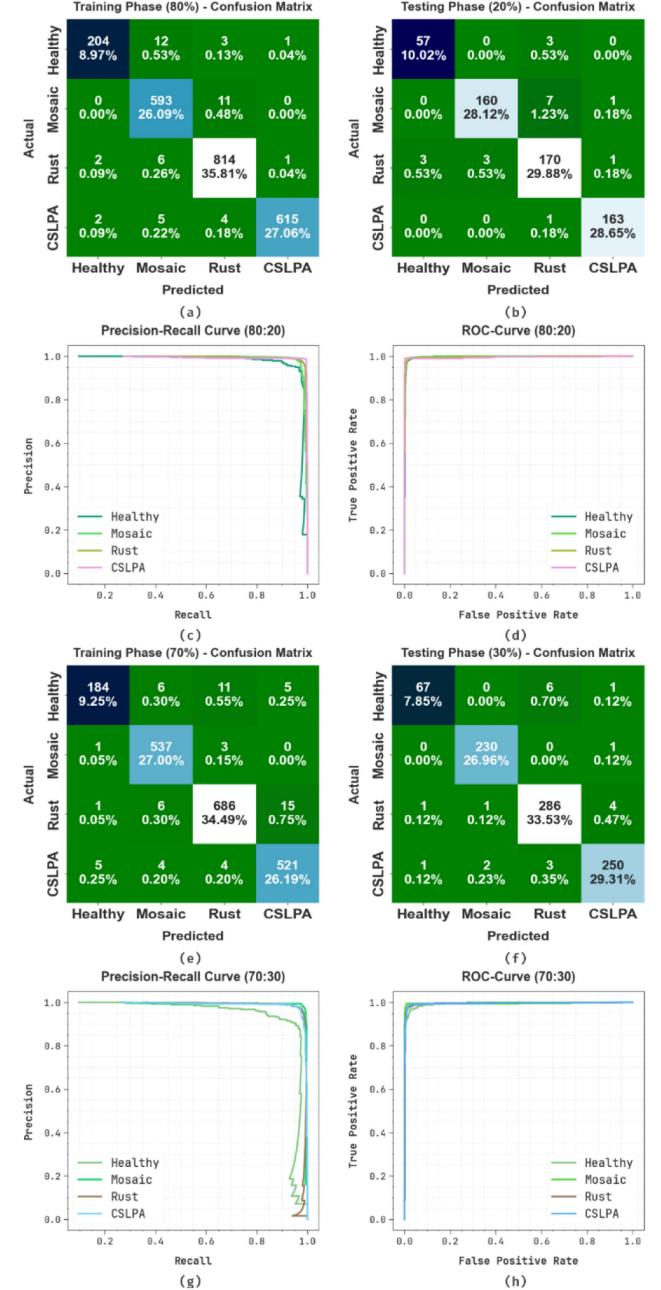


Fig. 5. 80:20 and 70:30 (a), (d), (e), (h) confusion matrices and (b), (c), (f), (g) curves of PR and ROC.

method is well-regularised and generalized. This reveals the method's stronger capability in learning and retaining beneficial features across both seen and unseen data.

Fig. 8 exemplifies the TRAIN and VALID losses of the RCHA-IFNNRSI method at 80:20 over 25 epochs. Initially, both TRAIN and VALID losses are higher, denoting that the process begins with a partial understanding of the data. As training evolves, both losses persistently decline, displaying that the method is efficiently learning and updating its parameters. The close alignment between the TRAIN and VALID loss curves during training indicates that the technique has not overfitted and upholds good generalization to unseen data.

TABLE III
CLASSIFIER OUTCOME OF THE RCHA-IFNNRSI APPROACH UNDER 80:20

Classes	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_l</i>	<i>F_{Measure}</i>	<i>MCC</i>
TRPHE (80%)					
Healthy	99.12	98.08	92.73	95.33	94.89
Mosaic	98.50	96.27	98.18	97.21	96.20
Rust	98.81	97.84	98.91	98.37	97.44
CSLPA	99.43	99.68	98.24	98.95	98.57
Average	98.97	97.96	97.01	97.47	96.77
TSPHE (20%)					
Healthy	98.95	95.00	95.00	95.00	94.41
Mosaic	98.07	98.16	95.24	96.68	95.34
Rust	96.84	93.92	96.05	94.97	92.68
CSLPA	99.47	98.79	99.39	99.09	98.72
Average	98.33	96.47	96.42	96.43	95.29

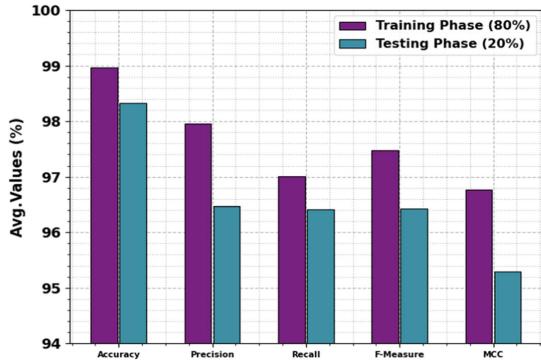


Fig. 6. Average values of the RCHA-IFNNRSI approach under 80:20.

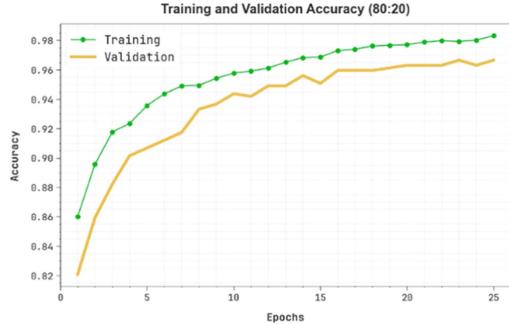


Fig. 7. Accu_y curve of RCHA-IFNNRSI model under 80:20.

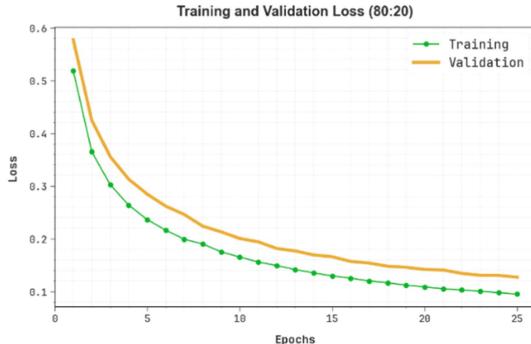


Fig. 8. Loss curve of the RCHA-IFNNRSI model under 80:20.

TABLE IV
CLASSIFIER OUTCOME OF THE RCHA-IFNNRSI MODEL UNDER 70:30

Classes	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_l</i>	<i>F_{Measure}</i>	<i>MCC</i>
TRPHE (70%)					
Healthy	98.54	96.34	89.32	92.70	91.97
Mosaic	98.99	97.11	99.26	98.17	97.49
Rust	97.99	97.44	96.89	97.17	95.61
CSLPA	98.34	96.30	97.57	96.93	95.80
Average	98.47	96.80	95.76	96.24	95.22
TSPHE (30%)					
Healthy	98.94	97.10	90.54	93.71	93.20
Mosaic	99.53	98.71	99.57	99.14	98.82
Rust	98.24	96.95	97.95	97.44	96.11
CSLPA	98.59	97.66	97.66	97.66	96.65
Average	98.83	97.60	96.43	96.99	96.19

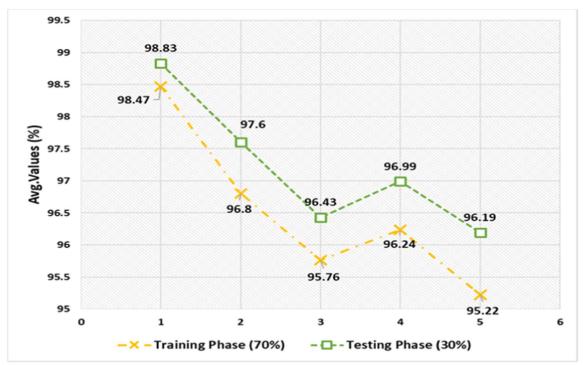


Fig. 9. Average values of the RCHA-IFNNRSI model under 70:30.

Table IV and Fig. 9 depict the classifier outcome of the RCHA-IFNNRSI approach at 70:30. On 70% TRPHE, the RCHA-IFNNRSI model attains an *accu_y*, *prec_n*, *recal_l*, *F_{Measure}*, and *MCC* of 98.47%, 96.80%, 95.76%, 96.24%, and 95.22%, respectively. At 30% TSPHE, the RCHA-IFNNRSI model attains an *accu_y*, *prec_n*, *recal_l*, *F_{Measure}*, and *MCC* of 98.83%, 97.60%, 96.43%, 96.99%, and 96.19%, respectively.

Fig. 10 exemplifies the training (TRAIN) *accu_y* and validation (VALID) *accu_y* of the RCHA-IFNNRSI technique at 70:30 over 25 epochs. In the beginning, both TRAIN and VALID *accu_y* rise rapidly, signifying effective pattern learning from the data. Around the epoch, the VALID *accu_y* minimally exceeds the training accuracy, signifying good generalization without overfitting. As training advances, it reflects maximum performance and a minimum performance gap between TRAIN and VALID. The close alignment of both curves in training denotes that the model is well-regularised and generalized. This exhibits the model's strong potential in learning and retaining valuable features across both seen and unseen data.

Fig. 11 exemplifies the TRAIN and VALID losses of the RCHA-IFNNRSI methodology on 70:30 over 25 epochs. Initially, both TRAIN and VALID losses are high, specifying that the model begins with a partial understanding of the data. As training progresses, both losses persistently reduce, indicating

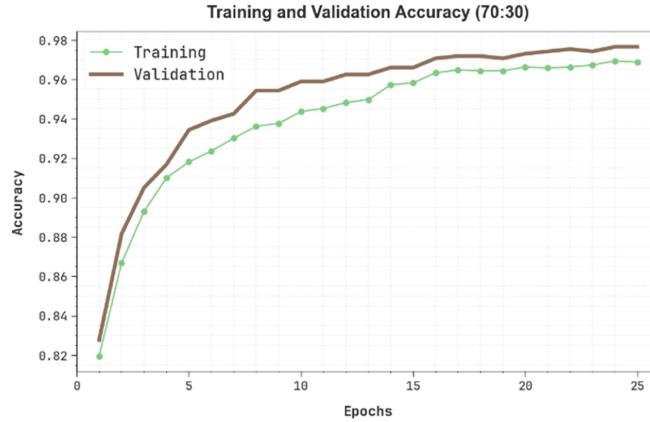
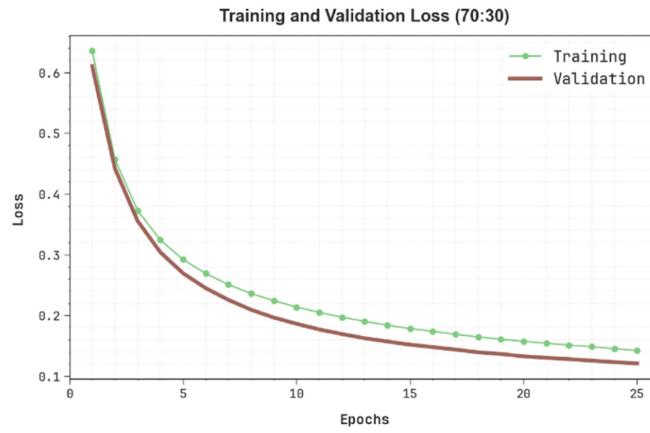
Fig. 10. Accu_y curve of RCHA-IFNNRSI technique under 70:30.

Fig. 11. Loss curve of RCHA-IFNNRSI technique under 70:30.

that the model is efficiently learning and enhancing its parameters. The close alignment between the TRAIN and VALID loss curves during training suggests that the model has not overfitted and retains good generalization to unseen data.

C. Discussion

The comparative study of the RCHA-IFNNRSI approach with existing models is illustrated in Table V and Fig. 12 [21], [22], [38], [39], [40]. The RCHA-IFNNRSI model achieved higher performance with accu_y of 98.97%, prec_n of 97.96%, recal_l of 97.01%, and F_{Measure} of 97.47%. At the same time, the present methodologies, such as YOLOv3, Faster RCNN, AgI-GAN, Inception-v3, DCNN, combined CNN-STARGAN, Plantention (MobileNetV2), GoogleNet, multi-SVM, and LDI-NET got lower values with accu_y of 93.15%, 85.20%, 96.29%, 92.50%, 84.60%, 95.60%, 98.34%, 94.76%, 97.80%, and 98.68%, respectively.

In Table VI and Fig. 13, the running time (RT) of the RCHA-IFNNRSI method with the present models is proven. Based on RT, the RCHA-IFNNRSI model offers the highest value of 98.97 s. At the same time, the YOLOv3, Faster RCNN, AgI-GAN, Inception-v3, DCNN, Combined CNN-STARGAN,

TABLE V
COMPARATIVE ANALYSIS OF THE RCHA-IFNNRSI MODEL WITH EXISTING TECHNIQUES

Model	Accu _y	Prec _n	Recal _l	F _{Measure}
YOLOv3	93.15	89.45	93.97	95.87
Faster RCNN	85.20	94.07	95.88	87.83
AgI-GAN	96.29	96.19	94.93	96.21
Inception-v3	92.50	88.88	93.29	95.24
DCNN	84.60	93.29	95.18	87.13
Combined CNN-STARGAN	95.60	97.62	94.36	96.66
Plantention (MobileNetV2)	98.34	93.19	97.19	96.85
GoogleNet	94.76	94.85	91.89	92.76
multi-SVM	97.80	88.45	96.69	97.10
LDI-NET	98.68	97.05	95.55	93.54
RCHA-IFNNRSI	98.97	97.96	97.01	97.47

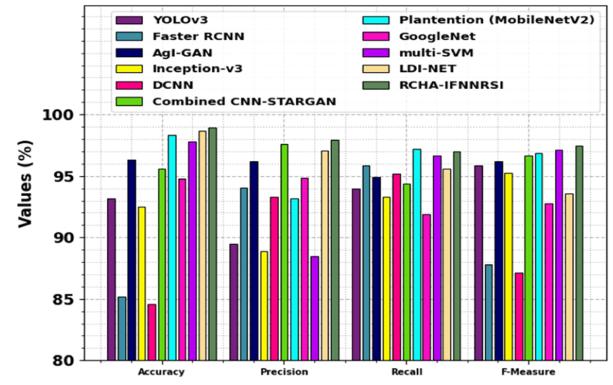


Fig. 12. Comparative analysis of the RCHA-IFNNRSI model with existing techniques.

TABLE VI
RT OUTCOME OF THE RCHA-IFNNRSI MODEL WITH EXISTING APPROACHES

Technique	RT (s)
YOLOv3	16.89
Faster RCNN	10.45
AgI-GAN	20.33
Inception-v3	92.50
DCNN	84.60
Combined CNN-STARGAN	95.60
Plantention (MobileNetV2)	98.34
GoogleNet	94.76
multi-SVM	97.80
LDI-NET	98.68
RCHA-IFNNRSI	98.97

Plantention (MobileNetV2), GoogleNet, multi-SVM, and LDI-NET methodologies got the least RT of 16.89, 10.45, 20.33, 92.50, 84.60, 95.60, 98.34, 94.76, 97.80, and 98.68 s, respectively.

Table VII and Fig. 14 specify the ablation study analysis of the RCHA-IFNNRSI methodology. The RCHA-IFNNRSI methodology demonstrated superior accu_y of 98.97%, prec_n of

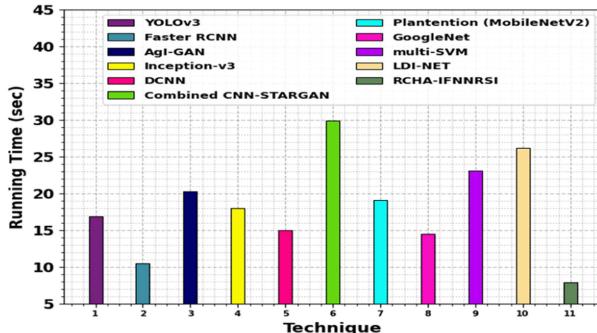


Fig. 13. RT outcome of RCHA-IFNNRSI model with existing approaches.

TABLE VII
RESULT ANALYSIS OF THE ABLATION STUDY OF THE RCHA-IFNNRSI METHODOLOGY

Model	<i>Accu_y</i>	<i>Prec_n</i>	<i>Recal_l</i>	<i>F_{Measure}</i>
ST-ResNet50	97.38	96.66	95.54	96.12
ACRNN	98.18	97.37	96.34	96.91
RCHA-IFNNRSI	98.97	97.96	97.01	97.47

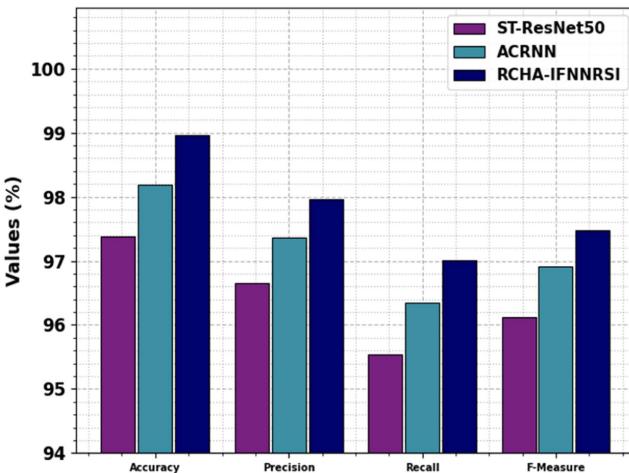


Fig. 14. Result analysis of the ablation study of the RCHA-IFNNRSI methodology.

97.96%, reca_l of 97.01%, and F_{Measure} of 97.47%. In comparison, ACRNN achieved an accu_y of 98.18%, prec_n of 97.37%, reca_l of 96.34%, and F_{Measure} of 96.91%, while ST-ResNet50 attained an accu_y of 97.38%, prec_n of 96.66%, reca_l of 95.54%, and F_{Measure} of 96.12%. These outputs confirm that the RCHA-IFNNRSI model illustrated robust performance across all key evaluation metrics.

V. CONCLUSION

In this article, the RCHA-IFNNRSI model is proposed for a crop health assessment framework utilizing RS and leaf image analysis for improving accuracy and decision-making in smart agriculture. Initially, the image preprocessing step includes BF-based noise removal and CLAHE-based contrast enhancement

to enhance the raw images by eliminating noise. For the feature extraction process, the model utilizes two generally used backbones, ST and ResNet-50. Ultimately, the ACRNN model is applied to identify and classify crops into distinct health status classes automatically. A wide range of experimentation of the RCHA-IFNNRSI technique is performed under the MH-SoyaHealthVision dataset. The comparison study of the RCHA-IFNNRSI technique demonstrated a superior accuracy value of 98.97% over existing models. The limitations of the RCHA-IFNNRSI technique comprise the relatively small and specific dataset used, which may restrict the generalizability of the results across diverse crop types and environmental conditions. Furthermore, the performance lacks validation on extensive field trials, which are crucial for evaluating real-world applicability. The computational complexity poses challenges for deployment on resource-constrained devices commonly used in agricultural settings. Future work should focus on expanding the dataset to include a wider variety of crops and geographical locations, and optimizing the model for faster inference and lower resource consumption. Integrating real-time data streaming and integration with existing farm management systems would further enhance practical adoption. Emphasizing scalability and user-friendly interfaces will be crucial to ensure farmers can effectively utilize the technology.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at <https://data.mendeley.com/datasets/hkbgh5s3b71/>, reference number [36].

REFERENCES

- [1] U. Shafi et al., “A multi-modal approach for crop health mapping using low altitude remote sensing, internet of things (IoT) and machine learning,” *IEEE Access*, vol. 8, pp. 112708–112724, 2020.
- [2] K. Ennouri and A. Kallel, “Remote sensing: An advanced technique for crop condition assessment,” *Math. Problems Eng.*, vol. 2019, no. 1, 2019, Art. no. 9404565.
- [3] S. Laveglia, G. Altieri, F. Genovese, A. Matera, and G. C. Di Renzo, “Advances in sustainable crop management: Integrating precision agriculture and proximal sensing,” *AgriEngineering*, vol. 6, no. 3, pp. 3084–3120, 2024.
- [4] R. Shukla et al., “Detecting crop health using machine learning techniques in smart agriculture system,” *J. Sci. Ind. Res.*, vol. 80, no. 08, pp. 699–706, 2021.
- [5] J. Wang, Y. Wang, G. Li, and Z. Qi, “Integration of remote sensing and machine learning for precision agriculture: A comprehensive perspective on applications,” *Agronomy*, vol. 14, no. 9, 2024, Art. no. 1975.
- [6] T. Fahey et al., “Active and passive electro-optical sensors for health assessment in food crops,” *Sensors*, vol. 21, no. 1, 2020, Art. no. 171.
- [7] A. I. Gálvez-Gutiérrez, F. Afonso, and J. M. Martínez-Heredia, “On the usage of deep learning techniques for unmanned aerial vehicle-based citrus crop health assessment,” *Remote Sens.*, vol. 17, no. 13, 2025, Art. no. 2253.
- [8] Y. Inoue, “Satellite-and drone-based remote sensing of crops and soils for smart farming—a review,” *Soil Sci. Plant Nutr.*, vol. 66, no. 6, pp. 798–810, 2020.
- [9] W. H. Maes and K. Steppe, “Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture,” *Trends Plant Sci.*, vol. 24, no. 2, pp. 152–164, 2019.
- [10] N. R. Dalezios, N. Dercas, N. V. Spyropoulos, and E. Psomiadis, “Remotely sensed methodologies for crop water availability and requirements in precision farming of vulnerable agriculture,” *Water Resour. Manage.*, vol. 33, no. 4, pp. 1499–1519, 2019.

- [11] R. Ajith Krishna, A. Kumar, and K. Vijay, "An automated optimize utilization of water and crop monitoring in agriculture using IoT," *J. Cogn. Hum.-Comput. Interaction*, vol. 1, no. 1, pp. 37–45, 2021.
- [12] H. Yang, P. Xu, S. Zhang, H. Kim, and I. Shin, "Construction of an intelligent analysis system for crop health status based on drone remote sensing data and CNN," *IEEE Access*, vol. 13, pp. 31643–31657, 2025.
- [13] A. S. Ibrahim et al., "AI-IoT based smart agriculture pivot for plant diseases detection and treatment," *Sci. Rep.*, vol. 15, no. 1, 2025, Art. no. 16576.
- [14] C. J. Young, H. J. Williams, E. F. Wardour, C. F. Ferro, E. N. Pask, and S. J. Lee, "Advanced crop management and health assessment using multispectral UAVs for precision agriculture," in *Proc. AIAA Aviation Forum Ascend*, 2025, Paper 3271.
- [15] A. L. P. De Ocampo and F. J. P. Montalbo, "A multi-vision monitoring framework for simultaneous real-time unmanned aerial monitoring of farmer activity and crop health," *Smart Agricultural Technol.*, vol. 8, 2024, Art. no. 100466.
- [16] R. K. Munaganuri and Y. N. Rao, "PAMICRM: Improving precision agriculture through multimodal image analysis for crop water requirement estimation using multidomain remote sensing data samples," *IEEE Access*, vol. 12, pp. 52815–52836, 2024.
- [17] S. K. Swarnkar, L. Dewangan, O. Dewangan, T. M. Prajapati, and F. Rabbi, "AI-enabled crop health monitoring and nutrient management in smart agriculture," in *Proc. 6th Int. Conf. Contemporary Comput. Informat.*, Sep. 2023, vol. 6, pp. 2679–2683.
- [18] Q. M. Ilyas, M. Ahmad, and A. Mahmood, "Automated estimation of crop yield using artificial intelligence and remote sensing technologies," *Bioengineering*, vol. 10, no. 2, 2023, Art. no. 125.
- [19] G. Venkatakotireddy, C. L. Reddy, J. Prabhakaran, and T. Nivethitha, "Development of high-quality crops using optimized machine learning in smart agriculture environment," in *Proc. 3rd Int. Conf. Artif. Intell. Smart Energy*, Feb. 2023, pp. 532–536.
- [20] A. Salam, Q. Javaid, and M. Ahmad, "Bio-inspired cluster-based optimal target identification using multiple unmanned aerial vehicles in smart precision agriculture," *Int. J. Distrib. Sensor Netw.*, vol. 17, no. 7, 2021, Art. no. 15501477211034071.
- [21] C. L. Fan and Y. J. Chung, "Integrating Image processing technology and deep learning to identify crops in UAV orthoimages," *Comput., Mater Continua*, vol. 82, no. 2, pp. 1925–1945, 2025.
- [22] C. F. Lacerda, Y. Ampatzidis, A. D. O. C. Neto, and V. Partel, "Cost-efficient high-resolution monitoring for specialty crops using AgI-GAN and AI-driven analytics," *Comput. Electron. Agriculture*, vol. 237, 2025, Art. no. 110678.
- [23] S. M. MirhoseiniNejad, D. Abbasi-Moghadam, and A. Sharifi, "ConvLSTM-ViT: A deep neural network for crop yield prediction using Earth observations and remotely sensed data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 17489–17502, 2024.
- [24] M. Zhang, C. Liu, Z. Li, and B. Yin, "From Convolutional networks to vision transformers: Evolution of deep learning in agricultural pest and disease identification," *Agronomy*, vol. 15, no. 5, 2025, Art. no. 1079.
- [25] N. Farmonov, M. Esmaeili, D. Abbasi-Moghadam, A. Sharifi, K. Amankulova, and L. Mucsi, "HypsLiDNet: 3-D–2-D CNN model and spatial–spectral morphological attention for crop classification with DE-SIS and LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 11969–11996, 2024.
- [26] U. S. Kumar et al., "Fusion of MobileNet and GRU: Enhancing remote sensing applications for sustainable agriculture and food security," *Remote Sens. Earth Syst. Sci.*, vol. 8, no. 1, pp. 118–131, 2025.
- [27] J. Mahmoodi, D. Abbasi-Moghadam, A. Sharifi, H. Nezamabadi-Pour, M. Esmaeili, and A. Vafaeinejad, "DESSA-net model: Hyperspectral image classification using an entropy filter with spatial and spectral attention modules on deepnet," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 14588–14613, 2024.
- [28] N. Hajare and A. S. Rajawat, "Black gram disease classification via deep ensemble model with optimal training," *Int. J. Image Graph.*, vol. 25, no. 04, 2025, Art. no. 2550033.
- [29] S. Wang, Y. Wang, J. Yue, H. Liang, Z. Zhang, and B. Li, "UniU-Net: A Unified U-Net deep learning approach for high-precision areca palm segmentation in remote sensing imagery," *Appl. Sci.*, vol. 15, no. 9, 2025, Art. no. 4813.
- [30] L. L. Scientific, "VGG-DAGSVM for multi-class apple leaf disease detection: A transfer learning approach," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 13, 2025.
- [31] S. Mumtaz et al., "Advanced leaf classification using multi-layer perceptron for smart crop management," *IEEE Access*, vol. 13, pp. 105579–105589, 2025.
- [32] B. Dong, Z. Wang, C. Chen, K. Wang, and J. Zhang, "An improved backbone fusion neural network for orchard extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 17961–17974, 2025.
- [33] M. Asif et al., "A novel hybrid deep learning approach for super-resolution and objects detection in remote sensing," *Sci. Rep.*, vol. 15, no. 1, 2025, Art. no. 17221.
- [34] G. Owino, T. Kamanu, J. Ndiritu, and C. B. Kikechi, "Adaptive infant cry classification using multi-armed bandit modality selection in an attentive convolutional recurrent neural network model," *Complex Intell. Syst.*, vol. 11, no. 9, 2025, Art. no. 375.
- [35] A. Salam et al., "Efficient prediction of anticancer peptides through deep learning," *PeerJ. Comput. Sci.*, vol. 10, 2024, Art. no. e2171.
- [36] [Online]. Available: <https://data.mendeley.com/datasets/hkbgh5s3b7/1>
- [37] S. Shinde and V. Attar, "An Indian UAV and leaf image dataset for integrated crop health assessment of soybean crop," *Data Brief*, vol. 60, 2025, Art. no. 111517.
- [38] S. T. Y. Ramadan et al., "Improving wheat leaf disease classification: Evaluating augmentation strategies and CNN-based models with limited dataset," *IEEE Access*, vol. 12, pp. 69853–69874, 2024.
- [39] B. Subburaj, M. Rohan, S. Ananthanarayanan, and D. Won, "Plantention: A general-purpose, lightweight and attention-based model for multi-crop leaf disease classification," *Results Eng.*, vol. 26, 2025, Art. no. 105075.
- [40] S. K. Ray et al., "Enhanced plant health monitoring with dual head CNN for leaf classification and disease identification," *J. Agriculture Food Res.*, vol. 21, 2025, Art. no. 101930, doi: [10.1016/j.jafr.2025.101930](https://doi.org/10.1016/j.jafr.2025.101930).