LAGRANGE CONDITIONAL GENERATIVE ADVERSARIAL NETWORK (LCGAN) BASED IMAGE DENOISING AND MULTI-SCALE DUAL ATTENTION INCEPTION V3 (MSDAIV3) FOR TOMATO LEAF DISEASES DETECTION

[1]N.S. Tamil Ilakkiya, [2]Dr.P.M. Gomathi

[1]Ph.D, Research scholar, Department of Computer science,

PKR Arts College for Women, Gobi, Bharthiyar University, Coimbatore, Tamil Nadu, India

[2] Associate Professor, Dean and Supervisor, Department of Computer Science,

PKR Arts College for Women, Gobi, Bharthiyar University, Coimbatore, Tamil Nadu, India

ABSTRACT: Agriculture has been the primary source of income for the majority of people in India. The impact of plant illness ranges from mild manifestation to the destruction of complete plantations that severely affect the agricultural economy. Therefore, early detection and diagnosis of these diseases are essential. During diagnosis, noise can be easily generated during image acquisition, transmission, and processing, which makes it challenging to extract disease features and reduces diagnosis accuracy. In this paper, new identification model is introduced for tomato leaf disease. Firstly, a Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced to reduce the image noise interference. LCGAN, noise removed image is estimated by an end-to-end trainable neural network. LCGAN includes of encoder and decoder architecture so that it can generate better detection results. LCGAN is used to decrease the difficulty of extracting tomato leaf disease features in the identification network. Secondly, three-segment linear conversion is utilized to increase image contrast. Thirdly, Inception-V3 is proposed to extract abundant disease features. In Inception-v3, a batch normalization (BN) layer is inserted as a regularizer between the auxiliary classifier and the fully connected layer. Finally, Multi-Scale Dual Attention Inception V3 (MSDAIV3) model is introduced to measure the inter-class similarity and intra-class variability identification of tomato leaf diseases. The proposed model has been trained and tested extensively on real time dataset with six classes (Bacterial spot, early blight, Leaf Mold, Septoria leaf spot, Leaf Curl Virus, and Healthy). Denoising results of proposed model is compared to existing methods using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Feature SIMilarity Index for Color images (FSIMc). The proposed approach shows its superiority over the existing methods using metrics like precision, recall, f-measure, and accuracy.

INDEX TERMS: Image Denoising, Lagrange Conditional Generative Adversarial Network (LCGAN), Inception-V3, Deep Learning, Multi-Scale Dual Attention Inception V3 (MSDAIV3), Feature extraction, Classification, and tomato leaf diseases.

1. INTRODUCTION

According to a prediction by the Food and Agriculture Organization (FAO) of the United Nations, the world population could reach 9.1 billion by 2050, ultimately increasing food requirements. Food production needs to be increased currently to feed the expected population of 10 billion people on the earth [1,2]. Various natural and artificial activities significantly reduce food production while anticipating reaching the expected level. Food security has been increasingly addressed; many countries and institutions are working to increase food production. In general, plant diseases, and leaf diseases are a major threat to the modern agricultural industry, heavily reducing the production and food quality [3]. However, the presence of several plant diseases also causes a significant reduction in both food quality and quantity.

Tomato is one of the most economically and nutritionally essential vegetables, which is cultivated all around the globe. Tomato crops can be easily affected by numerous diseases that can cause dramatic economic losses and food shortages. Hence, farmers need to diagnose tomato diseases as early as possible to reduce the risk of losing yields. Apart from preventing tomato diseases by annually testing the garden soil and maintaining an adequate level of potassium, in the past farmers used manual inspection of affected leaves to identify tomato diseases [4].

The traditional method of detecting and classifying diseases by physical observation is not always reliable and may result in a significant decrease in agricultural production [5]. Plant diseases attack the leaf initially before infecting the entire plant, reducing production quality and quantity [6]. The diagnosis of tomato leaf disease involves three stages: image denoising, feature extraction, and image recognition. It has become an important task to remove noise from the image and restore a high-quality

image. Physical features, such as shape, texture, and color, are used to create feature vectors, and traditional machine-learning techniques are used to identify feature vectors. Machine learning methods to help identify plant leaf diseases. Therefore, it is necessary to use a specific image denoising algorithm have been widely applied in some specific fields.

Traditional crop disease diagnosis is time-intensive and it can yield inconsistent results. Traditional visual inspection often results in low efficiency and insufficient accuracy. Studies have shown that this approach can lead to misdiagnosis because of its reliance on the subjective decision of human experts, particularly when diseases present similar symptoms. Another major limitation is that traditional methods are labor-intensive and slow. Given the vast amount of crops that need to be inspected in large-scale agricultural operations, relying solely on human expertise can be impractical and costly [7]. Automated crop disease diagnosis is vital for improving agricultural productivity, especially with the increasing global population.

Machine Learning (ML) and Deep Learning (DL) methods address these challenges by improving diagnostic accuracy and speed, which are critical for efficient disease management and boosting agricultural yields. The application of ML and DL techniques offers several benefits, including higher crop yields, better disease management, improved food security, and reduced chemical use. These technologies support precision agriculture, agricultural research, crop insurance, farm management, and environmental monitoring. During disease detection, increasing generation of digital images captured in poor conditions, image denoising methods have become an imperative tool for computer-aided analysis. Nowadays, the process of restoring information from noisy images to obtain a clean image is a problem of urgent importance.

Image denoising procedures remove noise and restore a clean image. A major problem in image denoising is how to distinguish between noise, edge, and texture. DL techniques have made remarkable progress in this field in recent years. DL methods, with their powerful learning capability and adaptability, can learn complex relationships between noise and images from large amounts of data [8]. This enables deep learning methods to better understand image content and effectively remove noise. Compared to traditional image denoising models, deep learning-based image denoising approaches demonstrate higher flexibility and robustness when dealing with noise in real-world scenarios [9]. The advent and progressive development of deep neural networks have catalyzed substantial advancements in learning-based denoising methods [10-12], marking a significant evolution in this field.

In existing deep learning-based tomato leaf disease identification algorithms, there are two factors limit the performance: 1) noise can be easily generated during image acquisition, transmission, and processing, which makes it challenging to extract disease features; 2) inter-class similarity and intra-class variability of tomato leaf diseases make it challenging to identify disease images. In this paper, new identification model is introduced for tomato leaf disease. Firstly, a Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced to reduce the image noise interference and it is used to decrease the difficulty of extracting tomato leaf disease features in the identification network. Secondly, three-segment linear conversion is utilized to increase image contrast. Thirdly, Inception-V3 is proposed to extract abundant disease features. Finally, Multi-Scale Dual Attention Inception V3 (MSDAIV3) model is introduced to solve the problem of inter-class similarity and intra-class variability identification of tomato leaf diseases.

2. LITERATURE REVIEW

Chen et al., [13] proposed a new framework for tomato leaf disease recognition. Firstly, the image is denoised and enhanced by Binary Wavelet Transform combined with Retinex (BWTR), noise points and edge points are removed, and important texture information is retained. Then, the tomato leaves were separated from the background using KSW optimizatied by Artificial Bee Colony algorithm (ABCK). Finally, the Both-channel Residual Attention Network model (B-ARNet) was used to identify the images. The application results of 8616 images show that the overall detection accuracy of 89.00%. Experiments show that the tomato leaf disease recognition method based on the combination of ABCKBWTR and B-ARNet is effective.

Peng et al., [14] proposed a novel classification network for tomato leaf disease, the Dense Inception MobileNet-V2 parallel convolutional block attention module network (DIMPCNET). To begin, a total of 1256 original images of 5 tomato leaf diseases are collected and expanded them to 8190 using image enhancement techniques. Next, an improved bilateral filtering and threshold function (IBFTF) algorithm is designed to effectively remove noise. Then, the Dense Inception convolutional neural network (DI) was designed to alleviate the problem of large intra-class differences and small inter-class differences. Then, a parallel convolutional block attention module (PCBAM) was added to MobileNet-V2

to reduce the impact of complex backgrounds. Finally, the experimental results show that the recognition accuracy and F1-score obtained by DIMPCNET. This method is the most advanced and provides a new idea for the identification of crop diseases, such as tomatoes, and the development of smart agriculture.

Islam et al., [15] proposed a high-performance tomato leaf disease detection approach, namely attention-based dilated CNN logistic regression (ADCLR). In preprocessing, bilateral filtering is introduced to handle larger features to make the image smoother and the Ostu image segmentation process to remove noise in a fast and simple way. In this proposed method, preprocess the image with bilateral filtering and Otsu segmentation. Then, a new feature extraction method using attention-based dilated CNN is used to extract most relevant features in a faster time. Finally, Conditional Generative Adversarial Network (CGAN) model is used to generate a synthetic image from the image which is preprocessed in the previous stage. The synthetic image is generated to handle imbalance and noisy or wrongly labeled data to obtain good prediction results. Then, the extracted features are normalized to lower the dimensionality. Finally, extracted features from preprocessed data are combined and then classified using fast and simple Logistic Regression (LR). The experimental outcomes show the state-of-the-art performance on the Plant Village database of tomato leaf disease for multiclass. From the experimental analysis, it is clearly demonstrated that the proposed multimodal approach can be utilized to detect tomato leaf disease precisely, simply and quickly.

Zhou et al., [16] proposed a restructured residual dense network (RDN) for tomato leaf disease identification. This hybrid deep learning model combines the advantages of deep residual networks and dense networks, which can reduce the number of training process parameters to improve calculation accuracy as well as enhance the flow of information and gradients. RDN model was first used in image super resolution, so we need to restructure the network architecture for classification tasks through adjusted input image features and hyper parameters. Experimental results show that this model can achieve top-1 average identification accuracy on the Tomato test dataset in AI Challenger 2018 datasets, which verifies its satisfactory performance. The restructured RDN model can obtain significant improvements over most of the state-of-the-art models in crop leaf identification, as well as requiring less computation to achieve highest performance.

Zhang et al., [17] proposed Asymptotic Non-Local Means algorithm (ANLM) to reduce the image's noise interference and to decrease the difficulty of extracting tomato leaf disease features in the identification network. Then, a Multi-channel Automatic Orientation Recurrent Attention Network (M−AORANet) is proposed to extract abundant disease features. An automatic orientation attention network is designed to locate lesion sites on tomato leaves. The fine multiscale feature is extracted and recycled to solve the problem of inter-class similarity and intra-class variability identification of tomato leaf diseases. Experimental results on 7493 images demonstrated that the identification accuracy of M−AORANet, which outperformed other current identification networks in comparison experiments. It can effectively provide decision information for tomato disease identification systems in precision agriculture.

Zhang et al., [18] proposed a new data augmentation method based on deep threshold multi-feature extraction convolution GAN with Mixed Attention and Markovian Discriminator (MMDGAN) for tomato disease leaf classification. Firstly, in the generator of MMDGAN, a deep threshold multi-feature extraction module was proposed to improve the feature extraction of tomato disease leaves. Then, a mixed attention mechanism combined cross attention module with fused features-highlighting module was proposed to coordinate the overall generation of images. Finally, Markov discriminator was used to strengthen the similarity judgment of local texture of images. PlantVillage dataset, the Frechet Inception Distance (FID) score of healthy tomato leaf image, Leaf Mold, Leaf Curl and Spider Mite generated by MMDGAN.

Patil and Manohar [19] developed an efficient tomato crop leaf disease segmentation model using an Enhanced Radial Basis Function Neural Network (ERBFNN). Initially, the noise present in the images is removed by a Gaussian filter followed by Contrast-Limited Adaptive Histogram Equalization (CLAHE) based on contrast enhancement and un-sharp masking. Then, color features are extracted from each leaf image and given to the segmentation stage to segment the disease portion of the input image. ERBFNN is enhanced using the modified sunflower optimization (MSFO) algorithm. ERBFNN approach is estimated using different metrics such as accuracy, Jaccard Coefficient (JC), Dice Coefficient (DC), precision, recall, F-Measure, sensitivity, specificity, and mean intersection over union (MIoU) and is compared with existing state-of-the-art methods of Radial Basis Function (RBF), Fuzzy C-Means (FCM), and Region Growing (RG).

Zhang et al., [20] proposed a method for classifying and recognizing tomato leaf diseases based on machine vision. First, to enhance the disease feature details in images, a linear transformation method is used for image enhancement, and oversampling is employed to expand the dataset, compensating for the imbalanced dataset. Next, a dual attention mechanism with convolutional (DAC) Block is used to construct a lightweight model named LDAMNet. The DAC block innovatively uses Hybrid Channel Attention (HCA) and Coordinate Attention (CSA) to process channel information and spatial information of input images respectively, enhancing the model's feature extraction capabilities. Additionally, Robust Cross-Entropy (RCE) loss function is robust to noisy labels, aimed at reducing the impact of noisy labels on the LDAMNet model during training.

Badiger and Mathew [21] presented DL technique for detecting and classifying the tomato disease and used Deep Batch-Normalized eLu Alexnet (DbneAlexnet) for classifying the tomato plant leaves. Initially, tomato plant leaf images are collected from specific dataset represented and it is subjected to preprocessing to eliminate unwanted distortions using anisotropic filtering. Then, the segmentation is carried out using U-net, which is trained by Gradient-Golden Search Optimization (Gradient-GSO) Algorithm and it is incorporation of both GSO and Gradient concept. Thereafter the segmented image is introduced to image augmentation process, where position augmentation and color augmentation. Finally, the multiclass plant leaf disease is classified using DbneAlexnet and is trained using proposed Gradient Jaya- Golden Search Optimization (GJ-GSO).

3. PROPOSED METHODOLOGY

In this paper, Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced to reduce the image noise interference. LCGAN, noise removed image is estimated by an end-to-end trainable neural network. Secondly, three-segment linear conversion is utilized to increase tomato leaf image contrast enhancement. Thirdly, Inception-V3 is proposed to extract abundant disease features. Finally, Multi-Scale Dual Attention Inception V3 (MSDAIV3) model is introduced to measure the inter-class similarity and intra-class variability identification of tomato leaf diseases. The proposed model has been trained and tested extensively on real time dataset with six classes ((Bacterial spot, early blight, Leaf Mold, Septoria leaf spot, Leaf Curl Virus, and Healthy)). Overall process of proposed system is illustrated in figure 1.

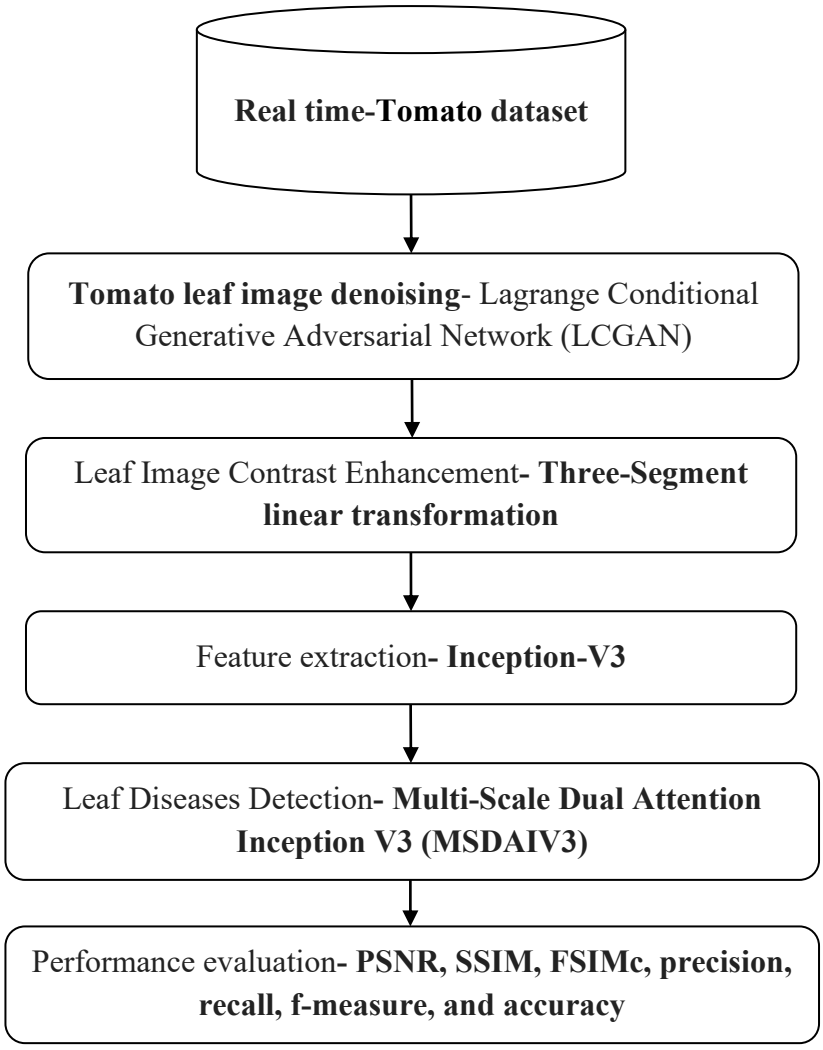

FIGURE 1. PROPOSED TOMATO LEAF DISEASES DETECTION FRAMEWORK

REFERENCES

3.1. DATASET COLLECTION

In this experiment, the tomato leaf dataset is collected from the real-time dataset. This public dataset is introduced for image processing and computer vision contains 9 classes of diseased and healthy plant leaves of tomato. In this work, a dataset of six classes including Bacterial spot, early blight, Leaf Mold, Septoria leaf spot, Leaf Curl Virus, and Healthy as shown in Figure 2. The image has a Red Green Blue (RGB) color model with dimensions of 256x256 pixels, so it resizes the image to 224x224 pixels and removes the image background before training the model. Totally 10239 images of tomato leaves were collected by the real time, including 1702 images of Bacterial spot tomato leaves, 800 images of early blight tomato leaves, 762 images of early mold tomato leaves, 1417 images of septoria leaf spot tomato leaves, 4286 images of Leaf Curl Virus tomato leaves and 1272 healthy leaf images, all of them are saved in jpg format.

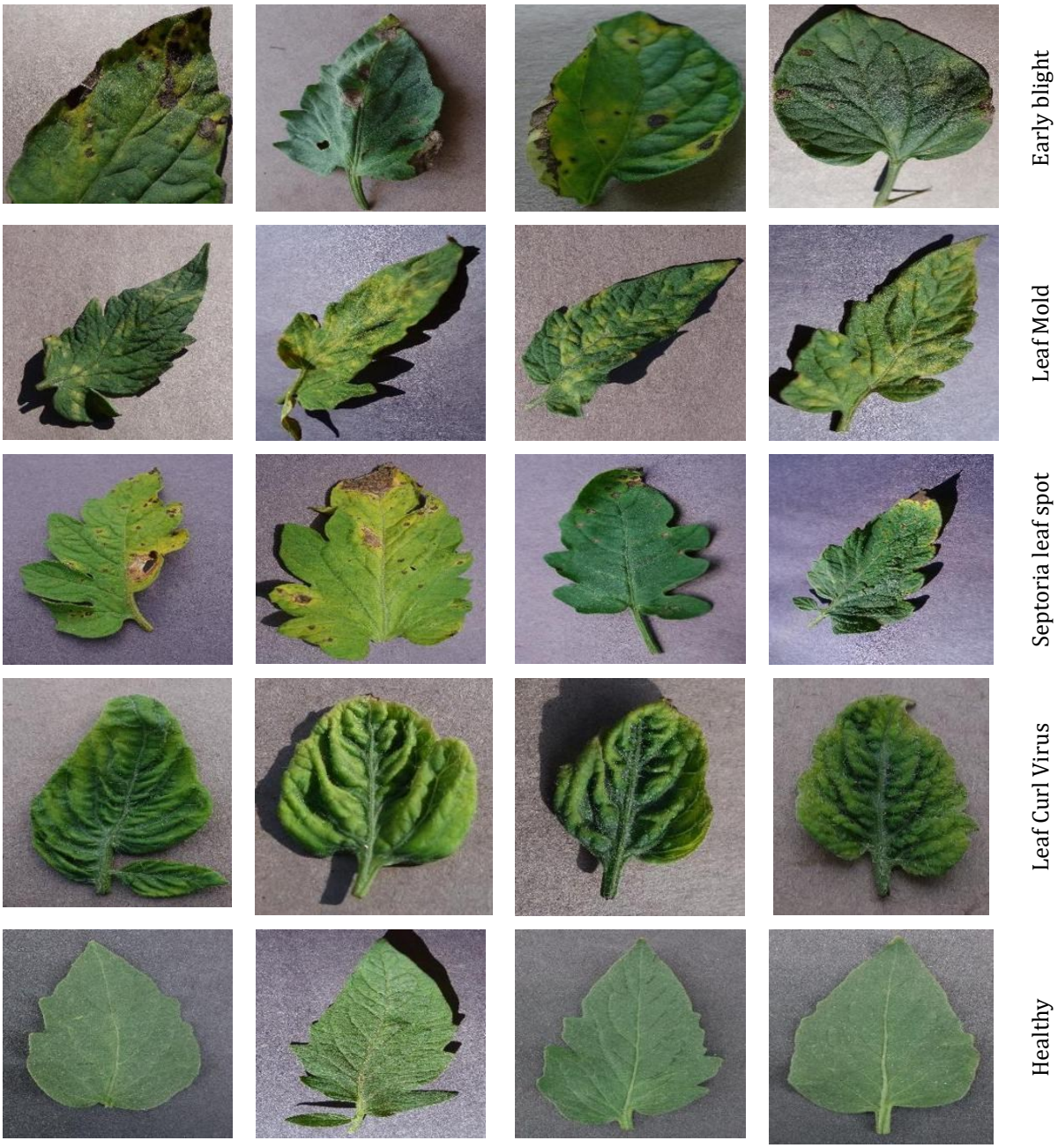

FIGURE 2. SAMPLE OF TOMATO LEAVES FROM THE REAL TIME DATASET

Table 1, and the dataset was divided into three folders: training 70%, Testing 20%, and validation 10%. The image of tomato leaves in the natural environment was collected by the camera, and then the image enhancement algorithm was used to de-noise and enrich the original image.

TABLE 1. TOMATO LEAF DISEASE DATASET

DISEASE TYPE	TOTAL IMAGES	TRAINING	TESTING	VALIDATION
Bacterial spot	1702	1191	340	170
Early blight	800	560	160	80
Leaf Mold	762	535	152	76
Septoria leaf spot	1417	992	283	142
Leaf Curl Virus	4286	3000	857	429
Healthy	1272	891	254	127

3.2. LAGRANGE CONDITIONAL GENERATIVE ADVERSARIAL NETWORK (LCGAN) BASED DENOISING

Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced by enforcing an additional constraint that the denoised image must be identical from its ground truth image. The adversarial loss from GAN provides additional lagrange regularization and helps to achieve superior results. The loss function is aimed at reducing noise artifacts introduced by GAN and ensure better visual quality. The generator sub-network is constructed using the introduced densely connected networks, whereas the multi-scale discriminator is proposed to leverage both local and global information to determine whether the corresponding denoised image is actual and noisy.

3.2.1. LCGAN architecture

LCGAN is used to directly learn a mapping from an input noisy image to a denoised image by constructing a conditional GAN. The proposed network is composed of three important parts (generator, discriminator and perceptual loss function) that serve distinct purposes. Similar to traditional GAN [22], the proposed method contains two sub-networks: a generator sub-network G and a discriminator sub-network D. The generator subnetwork G is a densely-connected symmetric deep CNN network with appropriate skip connections as shown in the top part in figure 3. Its primary goal is to synthesize a denoised image from an image that is degraded by noise. The multi-scale discriminator sub-network D as shown in the bottom part in Figure 3, serves to distinguish 'denoised image' (synthesized by the generator) from corresponding ground truth 'noise' image. It can also be viewed as guidance for the generator G. GAN are known to be unstable to train which results in artifacts in the output image synthesized by G, define a refined perceptual loss functions to address this issue. Additionally, this new refined loss function ensures that the generated denoised images are visually appealing.

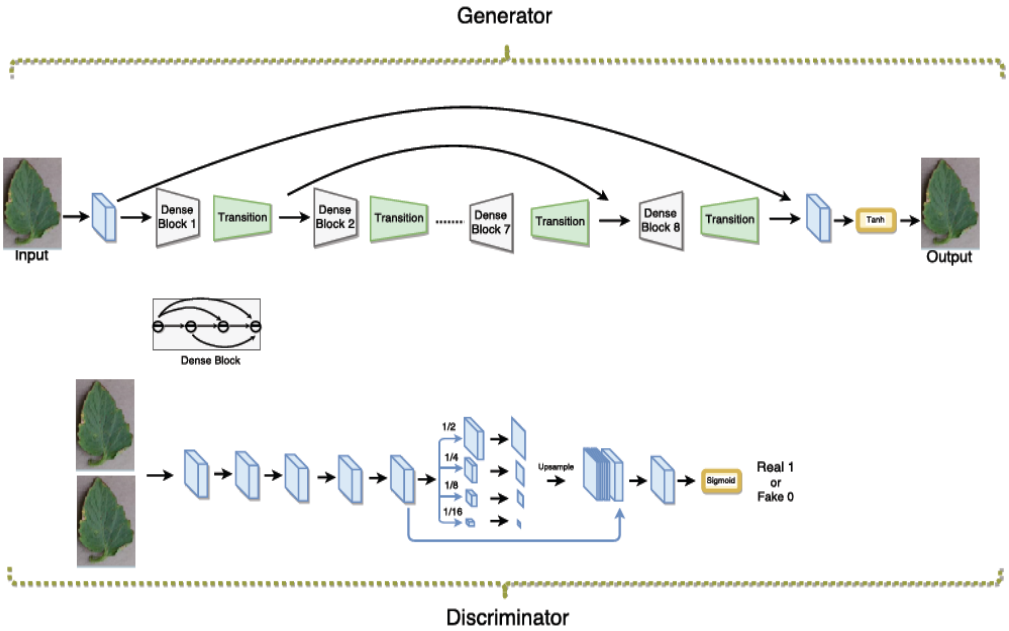

FIGURE 3. LAGRANGE CONDITIONAL GENERATIVE ADVERSARIAL NETWORKS (LCGAN) METHOD FOR IMAGE DENOISING

3.2.1.1. Generator with Symmetric Structure

As the goal of image denoising is to generate pixel level denoising image, the generator should be able to remove noises as much as possible without losing any detail information of the background image. So the key part lies in designing a good structure to generate denoising image. A symmetric structure is adopted to form the generator sub-network. The generator G directly learns an end-to-end mapping from input rainy image to its corresponding ground truth. Generators with densely connected blocks give more denoising results [23]. These dense blocks enable strong gradient flow and result in improved parameter efficiency. Furthermore, skip connections are introduced across the dense blocks to efficiently leverage features from different levels and guarantee better convergence. The j^{th} dense block D_j is represented as follows,

$$D_j = \mathrm{cat}\left[D_{j,1}, D_{j,2}, \ldots, D_{j,6}\right] \quad (1)$$

where $D_{j,i}$ represents the features from the i^{th} layer in dense block $D_j$and each layer in a dense block consists of three consecutive operations, Batch Normalization (BN), Leaky Rectified Linear Unit (LReLU) and a 3×3 convolution. Each dense block is followed by a transition block (T), functioning as up-sampling (T_u), down-sampling (T_d) or no-sampling operation (T_n). To make the network efficient in training and have better convergence performance, symmetric skip connections are included into the proposed generator subnetwork. Details of this architecture are discussed in table 2.

TABLE 2. NETWORK ARCHITECTURE FOR GENERATOR

Input, nc=3 (number of convolutions)
3×3 convolution, BN, ReLu, MaxP, nc=64
D(4 layers)+ T_d, nc=128
D(6 layers)+ T_d, nc=256
D(8 layers)+ T_n, nc=512
D(8 layers)+ T_n, nc=128
D(6 layers)+ T_u, nc =120
D(4 layers)+ T_u, nc =64
D(4 layers)+ T_u, nc=32
D(4 layers)+ T_u, nc=16
3×3 Convolution, Tanh, nc=3
Output, nc =3

3.2.1.2. Multi-scale Discriminator

GAN framework, the goal of denoising an input tomato image is not only to make the denoised result visually appealing and quantitatively comparable to the ground truth, but also to ensure that the denoised result is indistinguishable from the ground truth image. Therefore, a learned discriminator sub-network is designed to classify if each input image is actual/noisy. Similar to the structure [22], a convolutional layer with batch normalization and Parametric Rectified Linear Unit (PReLU) activation are used as a basis throughout the discriminator network. Then, a multi-scale pooling module, which pools features at different scales, is stacked at the end of the discriminator. The pooled features are then upsampled and concatenated, followed by a 1×1 convolution and a sigmoid function to produce a probability score normalized between [0,1]. By using features at different scales, explicitly incorporate global hierarchical context into the discriminator. The proposed discriminator sub-network D is shown in the bottom part of figure 3 and details of the multi-scale discriminator is shown in Table 3.

TABLE 3. NETWORK ARCHITECTURE FOR MULTI-SCALE DISCRIMINATOR

Input, nc=6
3×3 convolution, BN, ReLu, MaxP, nc=64
3×3 convolution, BN, ReLu, MaxP, nc=256
3×3 convolution, BN, ReLu, MaxP, nc=512
3×3 convolution, BN, ReLu, MaxP, nc=64
4 level pooling module, nc=72
Sigmoid activation function
Output, nc=72

3.2.1.3. Refined Perceptual Loss

New refined loss function is introduced to ensure that the results have good visual and quantitative scores along with good discriminatory performance [24]. Specifically, combine pixel-to-pixel Euclidean loss, perceptual loss and adversarial loss together with appropriate weights to form new refined loss function. The new loss function is then defined by equation (2) as follows,

$$L_{RP} = L_E * \lambda + \lambda_A L_A + \lambda_P L_P \tag{2}$$

where L_A represents adversarial loss from the D, $L_P$is perceptual loss and $L_E$is normal per-pixel loss function such as Euclidean loss, λ is the Lagrange multiplier between the original and noise removed image. Here, $\lambda_P$and λ_A are pre-defined weights for perceptual loss and adversarial loss. If set both $\lambda_P$and λ_Aa to be 0, then the network reduces to a normal CNN configuration, which aims to minimize only the Euclidean loss between denoised image and original image. If $\lambda_P$is set to 0, then the network reduces to a normal GAN. If λ_A set to 0, then the network reduces to the structure.

The three loss functions L_P, $L_E$and $L_A$are defined as follows. Given an image pair $\{x, y_b\}$ with C channels, width W and height H (i.e. $C \times W \times H$), where x is the input image and $y_b$is the corresponding ground truth, the per-pixel Euclidean loss is defined by equation (3) as follows,

$$L_E = \frac{1}{CWH}\sum_{c=1}^{C}\sum_{x=1}^{W}\sum_{y=1}^{H}\left\|\phi_E(x)^{c,w,h} - (y_b)^{c,w,h}\right\|_2^2 \tag{3}$$

where ϕ_E is the learned network G for generating the denoised output. Suppose the outputs of certain high-level layer are with size $C_i \times W_i \times H_i$. Similarly, the perceptual loss is defined by equation (4) as follows,

$$L_P = \frac{1}{C_i W_i H_i}\sum_{c=1}^{C}\sum_{x=1}^{W}\sum_{y=1}^{H}\left\|V(\phi_E(x))^{c,w,h} - V(y_b)^{c,w,h}\right\|_2^2 \tag{4}$$

where V represents a non-linear CNN transformation. Given a set of N denoised images generated from the generator $\{\phi_E(x)\}_{i=1}^{N}$, the entropy loss from the discriminator to guide the generator is defined by equation (5) as follows,

$$L_A = -\frac{1}{N}\sum_{i=1}^{N}\log(D(\phi_E(x))) \tag{5}$$

Once the images are denoised and then it is enhanced using contrast enhancement.

3.3. THREE-SEGMENT LINEAR TRANSFORMATION BASED CONTRAST ENHANCEMENT

Three-segment linear conversion is utilized to increase image contrast. Equation (6) is used for the three-segment linear transformation,

$$f(i,j) = \begin{cases} k_1 \times d(i,j), & 0 \le d(i,j) < a \\ b + k_2 \times d(i,j), & a \le d(i,j) \le c \\ d + k_3 \times d(i,j), & c < d(i,j) \le 255 \end{cases} \tag{6}$$

where $f(i,j)$ represents the final image after boosting, $d(i,j)$ is the input denoised image, k_1, k_2, and k_3 are the slopes of the three-segment transformation, and their expressions are as follows,

$$k_1 = \frac{a}{b}, k_2 = \frac{d-b}{c-a}, k_3 = \frac{255-d}{255-c} \tag{7}$$

(a, b) and (c, d) are the points on the function where the slope changes and the linear transformation raises the image contrast.

3.4. INCEPTION V3 BASED FEATURE EXTRACTION

InceptionV3 is introduced to find the best model as the backbone for the feature extractor [14]. Inception-v3 adopts convolutional kernels of different sizes, which enables it to own receptive fields of different areas. Inception-v3, a batch normalization (BN) layer is inserted as a regularizer between the auxiliary classifier and the fully connected (FC) layer. In the BN model, the batch gradient descent method can be employed to accelerate the training speed and model convergence of the deep neural network. These layers utilize compact filter sizes such as 1×1, 3×3, 1×3, and 3×1 to effectively reduce the number of trainable parameters. InceptionV3 operates on input images with a default size of 299×299×3. Initial image processing involves the application of five convolutional layers, each employing multiple 3×3 kernels. In the proposed framework, removed the final dense layers from the InceptionV3 model and obtain a feature vector with 8×8 and 2048 channels. The equations (8-13) of BN are expressed as follows,

$$B = \{\chi_{1,\dots m}\}, \gamma, \beta \tag{8}$$

$$\{y_i = BN_{\gamma,\beta}(\chi_i)\} \tag{9}$$

$$\mu_B \leftarrow \frac{1}{m}\sum_{i=1}^{m}\chi_i \tag{10}$$

$$\sigma_B^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m}(\chi_i - \mu_B)^2 \tag{11}$$

$$\hat{\chi}_i \leftarrow \frac{\chi_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{12}$$

$$y_i \leftarrow \gamma\hat{\chi}_i + \beta = BN_{\gamma,\beta}(\chi_i) \tag{13}$$

where x is the minimum activation value of batch B, m is the number of activation values, γ and β are learnable parameters (γ is responsible for adjusting the variance in the value distribution and β is responsible for adjusting the position of the average value), $\mu_B$represents the average value in one dimension, σ_B^2 is the standard deviation in each dimension of the feature map, and ε is a constant. Furthermore, in Inception-v3, large convolution kernels are divided into small convolution kernels in series, convolution and pooling are connected in parallel, and labels are added for regularization based on the smoothing criteria. In addition, considering the distribution inconsistency between inputs and outputs in a traditional deep neural network, which creates great obstacles for feature extraction, BN is introduced into Inception-v3. By normalizing the input into each layer, the learning effect is optimized by adam optimizer.

3.5. MSDAIV3 BASED LEAF DISEASES DETECTION

MSDAIV3 is introduced that integrates a deep learning model combining InceptionV3 with a modified multi-scale attention-based spatial attention block to enhance model performance. MSDAIV3 employs an InceptionV3 backbone with a fusion of dual attention modules to construct the proposed architecture. The InceptionV3 model generates rich features from images, capturing both local and global aspects, which are then enhanced by utilizing the modified multi-scale spatial attention block, resulting in a significantly improved feature map. MSDAIV3 is introduced that the integration of Channel Attention (CA) and Spatial Attention (SA). By leveraging this dual attention scheme, MSDAIV3 model can effectively focus on the leaf diseases classification. The proposed model, along with the CA and multi scale SA mechanisms, enables the most important part within the tomato images by extracting multi scale features resulting in improved performance. The incorporation of multi-scale SA enhances the ability of model is increased to capture most valuable features at different spatial resolutions, enabling a more comprehensive understanding of tomato images. Simultaneously, the CA module helps in prioritizing the most informative channels, further refining the feature representation. MSDAIV3 model is to accurately detect disease images with high accuracy compared to standard attention module. MSDAIV3 architecture achieves higher performance in disease classification by enabling the model to adaptively focus on the most discriminative features, resulting in improved accuracy and robustness for tomato leaf diseases.

3.5.1. Channel Attention (CA)

CA is introduced to utilize the inter-channel relationship of related features. The feature of each channel in the images is evaluated as a feature map. CA is shown in Figure 4 which consists of two different layers including a maximum convolution layers, pooling layer, global average pooling layer, and fully connected layers. During training, the importance of each convolution feature map of each channel does not represent disease detection, despite using either using max-pooling or average-pooling for different task. In this research, both the maximum-pooling and average-pooling techniques which can generate more enhanced feature maps.

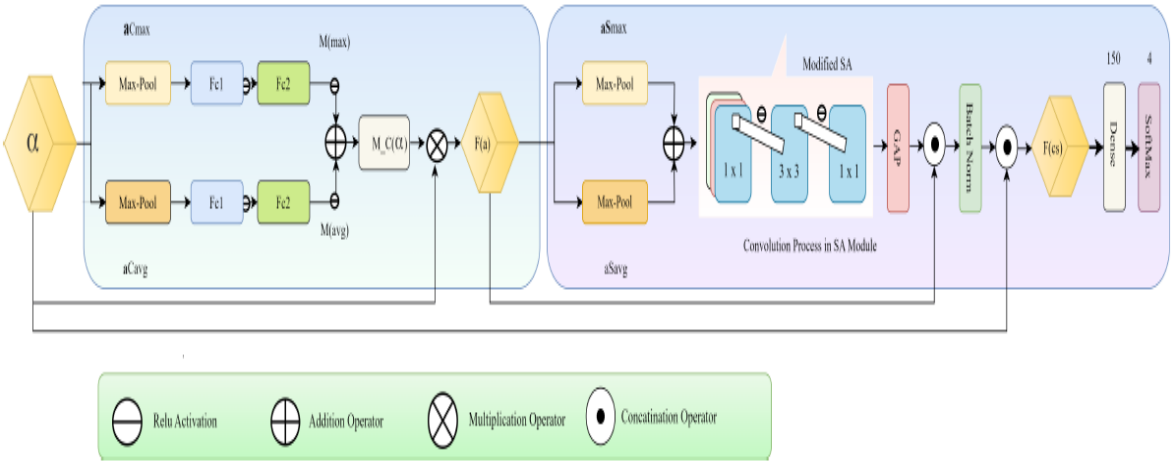

FIGURE 4. PROPOSED MULTI-SCALE MODEL ARCHITECTURE

Therefore, these distinct pooling strategies operation were performed separately on the spatial dimension of feature map α, as a result, two different spatial background descriptions: α_{Cavg} and α_{Cmax}, mathematically are calculated by using the equations(14-15),

$$\alpha_{Cavg} = Avg - Pool(\alpha) \tag{14}$$

$$\alpha_{Cmax} = Max - Pool(\alpha) \tag{15}$$

After these operation, two Fully Connected (FC) layers, fc^1 and fc^2, generate the output of $\alpha_{Cavg}$and α_{Cmax}, which shared common parameters. ReLU activation function is utilized on each FC layer to derive $M_{max}$and M_{avg}. By performing an addition operation on both feature maps, the weight coefficients Mc(α) are obtained as below,

$$M_{max} = R\left(fc^2\left(R(fc^1(\alpha_{Cmax}))\right)\right) \tag{16}$$

$$M_{avg} = R\left(fc^2\left(R\left(fc^1(\alpha_{Cavg})\right)\right)\right) \tag{17}$$

$$M_c(\alpha) = M_{max} \oplus M_{avg} \tag{18}$$

The fc^1 and fc^2layers employed 1 × 1 convolution operations to expand and compress the features of each channel. Here R is the ReLU activation function, ⊕is addition operation. Then the more concise features $M_c(\alpha)$by adding $M_{max}$and M_{avg}. To maintain the feature flow a skip connection, ⊗ is introduced resulting in the CA map F_c,

$$F_c = M_c(\alpha) \otimes \alpha \tag{19}$$

From this fully connected layer results are found by MSDAIV3 model.

3.5.2. Multi-Scale Spatial Attention (MSA)

MSA is developed a dual attention networks. Spatial attention maps are generated by analyzing the inter-spatial relationship within the features. The SA significantly focuses on the dominant elements of the image within spatial areas. This research consists of two different pooling operations, i.e., max pooling and average pooling. The outputs of these pooling layers are then integrated to generate an enhanced feature descriptor. These pooling operations give priority to critical regions within the features map, which results in the enhancement of the model generalization capability. Integrating max pooling and average pooling operations concurrently offers an efficient means of highlighting essential information embedded within the features. The mathematical expressions for computing max pooling and average pooling by equations (20-21),

$$\alpha_{Savg} = AveragePool(F_c) \tag{20}$$

$$\alpha_{Smax} = MaxPool(F_c) \tag{21}$$

The obtained feature vectors are then added and combined with an addition operator and the resultant feature maps are passed to the convolutional layer to generate a 2D SA feature map. Modified the SA module and utilized three convolutional layers with ReLU activation function. In the first layer,

utilizes the convolution layer of 1 × 1 with 64 filters, the second convolution layer uses a 3 × 3 with 64 layer, and the final layer uses a 1 × 1 convolution layer,

$$\mathrm{Ms}_{\mathrm{F_c}} = \mathrm{R}\left(\mathrm{f}^{1\times1}\left(\mathrm{R}\left(\mathrm{f}^{3\times3}\left(\mathrm{R}\left(\mathrm{f}^{1\times1}\left(\alpha_{\mathrm{Savg}} \oplus \alpha_{\mathrm{Smax}}\right)\right)\right)\right)\right)\right) \tag{22}$$

In the equation (22), f is the filter size. The SA map can be generated by utilizing global average pooling on the feature maps and then concatenating it with the $\mathrm{F_c}$ as shown in the equation (23),

$$\mathrm{Ms}_{\mathrm{F_{cGAP}}} = \mathrm{GAP}(\mathrm{F_c}) \tag{23}$$

$$\mathrm{F_s} = \left[\mathrm{Ms}_{\mathrm{F_{cGAP}}} \odot \mathrm{F_c}\right] \tag{24}$$

After the concatenation, batch normalization operation was performed on $\mathrm{F_s}$. Finally, these feature maps are concatenated from batch normalization and α to generate the $\mathrm{F_{CS}}$,

$$\mathrm{F_{CS}} = [\mathrm{Batchnorm} \odot \alpha] \tag{25}$$

Following the feature extraction technique from the $\mathrm{F_{CS}}$, the resultant features were then passed into a dense layer consisting of 150 neurons, using a soft max layer to assign input images to their appropriate classes. Subsequently, the output of the SA module traverses through three additional fully connected layers, each comprising 64, 32, and 2 neurons, respectively. The utilization of the SoftMax activation function facilitates the classification of input images based on their respective classes.

4. EXPERIMENTATION AND RESULTS

In these sections shows the performance comparison of detection methods, and image denoising methods using real time dataset. Dataset includes of six categories of tomato leaves, including Bacterial spot, Early blight, Leaf Mold, Septoria leaf spot, Leaf Curl Virus, and Healthy. The image of tomato leaves in the natural environment was collected by the camera, and then the image enhancement algorithm was used to de-noise and enrich the original image. The specific configuration of the experimental environment is shown in Table 4. MATLAR2024a was adopted as the deep learning framework. Table 4 displays the settings for the experimental parameters. The dimension of the input image is 224 × 224, the training procedure for the neural network is accelerated by GPU, and the adaptive moment estimator (Adam) algorithm is chosen as the model parameter optimizer. Cross-entropy loss is used as the loss function, the learning rate is 0.001, the neural network is trained with 60 pictures per batch (batch size), and the training period (epoch) is 200 epochs.

TABLE 4. EXPERIMENTATION SETUP

Configuration	Ranges
MATLAB	R2024a
CPU	Intel® Xeon® Gold 6271C CPU@ 2.60 GHz
GPU	NVIDIA Tesla V100 GPU32G
RAM	32 GB
Magnetic Disk	100 GB
Operating System	Windows 10(64 bits)
Image size	224×224
Batch Size	60
Learning rate	0.001
Epochs	200

4.1. EVALUATION INDEX

Denoising results of proposed model is compared to existing methods using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Feature SIMilarity Index for Color (FSIMc). The proposed disease detection approach shows its superiority over the existing methodologies using metrics like precision, recall, f-measure, and accuracy. The evaluation standards used in the experiment include subjective and objective evaluations. Subjective evaluation refers to visual inspection

of images, evaluating the denoising effect of the model. Objective evaluation uses peak signal to noise ratio (PSNR) and structural similarity (SSIM). PSNR is based on mean square error (MSE), is an image quality evaluation index. The higher the PSNR value, the better the image quality. In the experiment, the higher the PSNR value indicates a higher similarity between the denoised image and the original image. PSNR is calculated a follows,

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{2^{n}}{\mathrm{MSE}}\right)^{2} \tag{26}$$

$$\mathrm{MSE} = \frac{1}{H \times W}\sum_{j=1}^{H}\sum_{k=1}^{W}\left(M(j,k) - N(j,k)\right)^{2} \tag{27}$$

where M and N are the predicted and true values, respectively. j and k are all pixels in the image. H and W represent the height and width of the image, and n is set to 8. Structural Similarity (SSIM) is an evaluation metric to measure the similarity between two images. It estimates similarity based on image brightness, contrast, and structure. The mean is used as the brightness estimate, the standard deviation as the contrast estimate, and the covariance as the measure of structural similarity. SSIM is calculated as follows,

$$\mathrm{SSIM}(M,N) = \frac{(2\mu_m\mu_n + c_1)(2\sigma_{mn} + c_1)}{(\mu_m^2 + \mu_n^2 + c_1)(\sigma_m^2 + \sigma_n^2 + c_2)} \tag{29}$$

where $\mu_m$is the mean of image M, and $\mu_n$is the mean of image N. σ_m^2 and σ_n^2represent the variances of images M and N, respectively. $\sigma_{mn}$is denoted as the covariance between images M and N. $c_1$and $c_2$are constants for stability. The SSIM value ranges between 0 and 1, with a higher SSIM indicating more similarity. When SSIM is 1, the two images being compared are identical. Furthermore, we also utilized Feature Similarity Index for Color images (FSIMc), is quantitatively evaluate the score of image quality. In simple words, it can be used to measure the similarity between two color images. The FSIMc is an extension of the FSIM, which is originally designed for grayscale images,

$$\mathrm{FSIMc}(X,Y) = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\mathrm{PC}_m(x,y).\,\mathrm{GM}_m(x,y).\max\{S_l(x,y), S_r(x,y)\}}{\sum_{x=1}^{M}\sum_{y=1}^{N}\mathrm{PC}_m(x,y).\,\mathrm{GM}_m(x,y)} \tag{30}$$

where X and Y are the two color images being compared, $\mathrm{PC}_m$is the phase congruency at a given pixel, $\mathrm{GM}_m$is the gradient magnitude at the pixel, and $S_l$and $S_r$are the similarity measures for the left and right images respectively. The sums are taken over all pixels (x, y) in the images. The FSIMc score is a value between 0 and 1, where a higher value indicates greater similarity between the two images.

In this study, according to the confusion matrix, Precision, Recall, F-measure, and accuracy are chosen as assessment indicators to thoroughly assess the effectiveness of deep learning algorithms. False negative samples (FN), false positive samples (FP), true negative samples (TN), and true positive samples (TP) were used to create the assessment indexes (FN). These indicators are derived as follows,

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{31}$$

$$\mathrm{Recall} = \frac{TP}{TP + FN} \tag{32}$$

$$\mathrm{F-measure} = \frac{2TP}{2TP + FP + FN} \tag{33}$$

$$\mathrm{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{34}$$

$$T_A = \frac{T}{N} \tag{35}$$

where T is the overall detection time of the verification set, and N represents the verification set's overall detection time. In addition, the proposed model has the quickest average diagnostic time (T_A(ms)).

4.2. COMPARISON WITH TRADITIONAL DL METHODS

Table 5 shows the comparison of traditional denoising techniques with proposed model on both quantitative and qualitative evaluations with six classes. The quantitative assessment involved using important metrics such as PSNR, SSIM and FSIMc to numerically evaluate the quality of the denoised

images. Additionally, qualitative evaluation is performed by visually representing the restored images, allowing for an intuitive understanding of their visual quality and accuracy.

TABLE 5. QUALITY EVALUATION METRICS VS. IMAGE DENOISING METHODS

Methods/ Images	Bacterial spot			Early blight			Leaf Mold		
	PSNR (dB)	SSIM	FSIMc	PSNR (dB)	SSIM	FSIMc	PSNR (dB)	SSIM	FSIMc
BWTR	31.58	0.7741	0.8044	28.91	0.7557	0.7681	35.62	0.8361	0.8551
IBFTF	33.67	0.8212	0.8491	32.36	0.7891	0.8244	37.85	0.8675	0.8789
ANLM	36.41	0.8545	0.8719	35.65	0.8254	0.8558	39.74	0.8962	0.9054
LCGAN	40.72	0.9067	0.9158	38.58	0.8835	0.8983	42.69	0.9254	0.9348
Methods/ Images	Septoria leaf spot			Leaf Curl Virus			Healthy		
	PSNR (dB)	SSIM	FSIMc	PSNR (dB)	SSIM	FSIMc	PSNR (dB)	SSIM	FSIMc
BWTR	33.11	0.7877	0.8175	30.39	0.7641	0.7898	37.89	0.8497	0.8692
IBFTF	36.88	0.8364	0.8589	32.61	0.8069	0.8376	39.81	0.8781	0.8918
ANLM	38.23	0.8641	0.8878	35.58	0.8452	0.8682	42.66	0.9056	0.9289
LCGAN	41.94	0.9183	0.9296	38.72	0.8673	0.9057	45.75	0.9378	0.9418

TABLE 6. EVALUATION RESULTS VS.LEAF DISEASE DETECTION METHODS

Methods	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	Diagnostic time (ms)
B-ARNet	85.89	84.09	84.98	88.14	40.62
DIMPCNET	88.08	86.65	87.36	90.12	36.49
M–AORANet	89.57	88.63	89.09	91.63	33.75
MSDAIV3	91.76	91.19	91.47	93.52	30.47

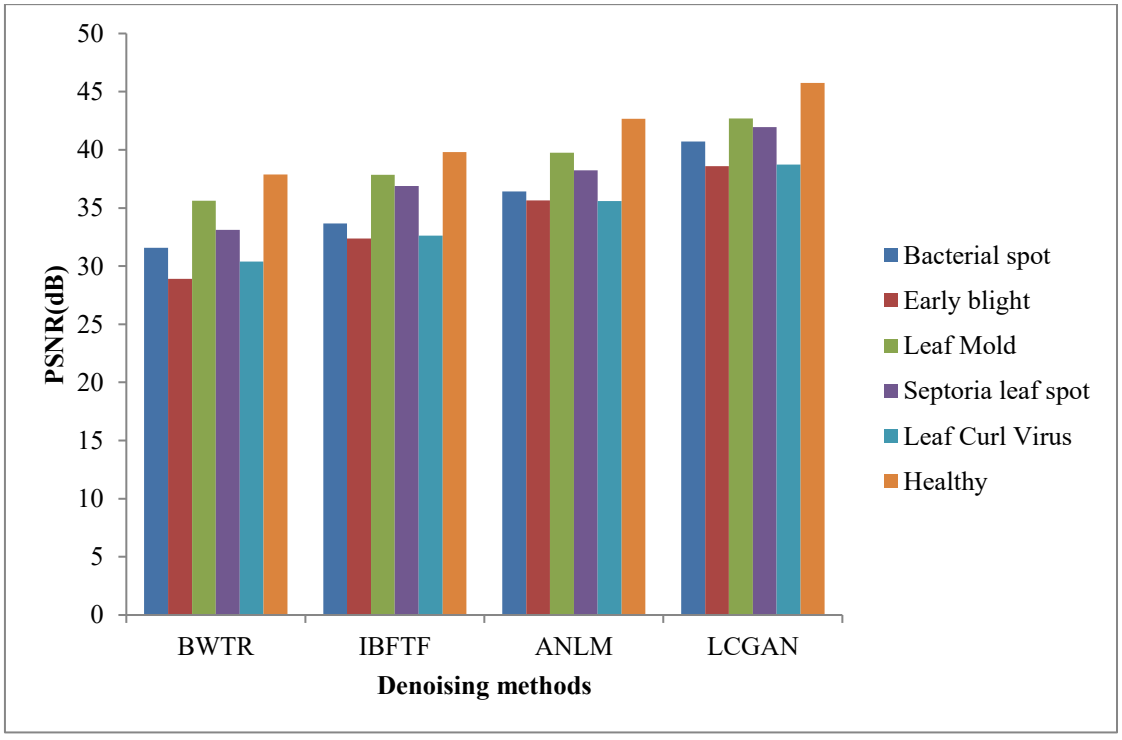

FIGURE 5. PSNR COMPARISON VS. DENOISING METHODS

PSNR comparison of denoising methods such as BWTR, IBFTF, ANLM, and LCGAN among bacterial spot, early blight, leaf mold, Septoria leaf spot, leaf curl virus, and healthy are illustrated in figure 5. LCGAN gives the highest PSNR of 40.72 dB, 38.58 dB, 42.69 dB, 41.94 dB, 38.72 dB, and 45.75 dB among bacterial spot, early blight, leaf mold, Septoria leaf spot, leaf curl virus, and healthy. BWTR, IBFTF, and ANLM give the lowest PSNR of 37.89 dB, 39.81 dB, and 42.66 dB for healthy. BWTR, IBFTF, and ANLM give the lowest PSNR of 31.58 dB, 33.67 dB, and 36.41 dB for bacterial spots. BWTR, IBFTF, and ANLM give the lowest PSNR of 28.91 dB, 32.36 dB, and 35.65 dB for early blight. BWTR, IBFTF, and ANLM give the lowest PSNR of 35.62 dB, 37.85 dB, and 39.74 dB for leaf mold. BWTR, IBFTF, and ANLM give the lowest PSNR of 33.11 dB, 36.88 dB, and 38.23 dB for septoria leaf spot. BWTR, IBFTF, and ANLM give the lowest PSNR of 30.39 dB, 32.61 dB, and 35.58 dB for the leaf curl virus.

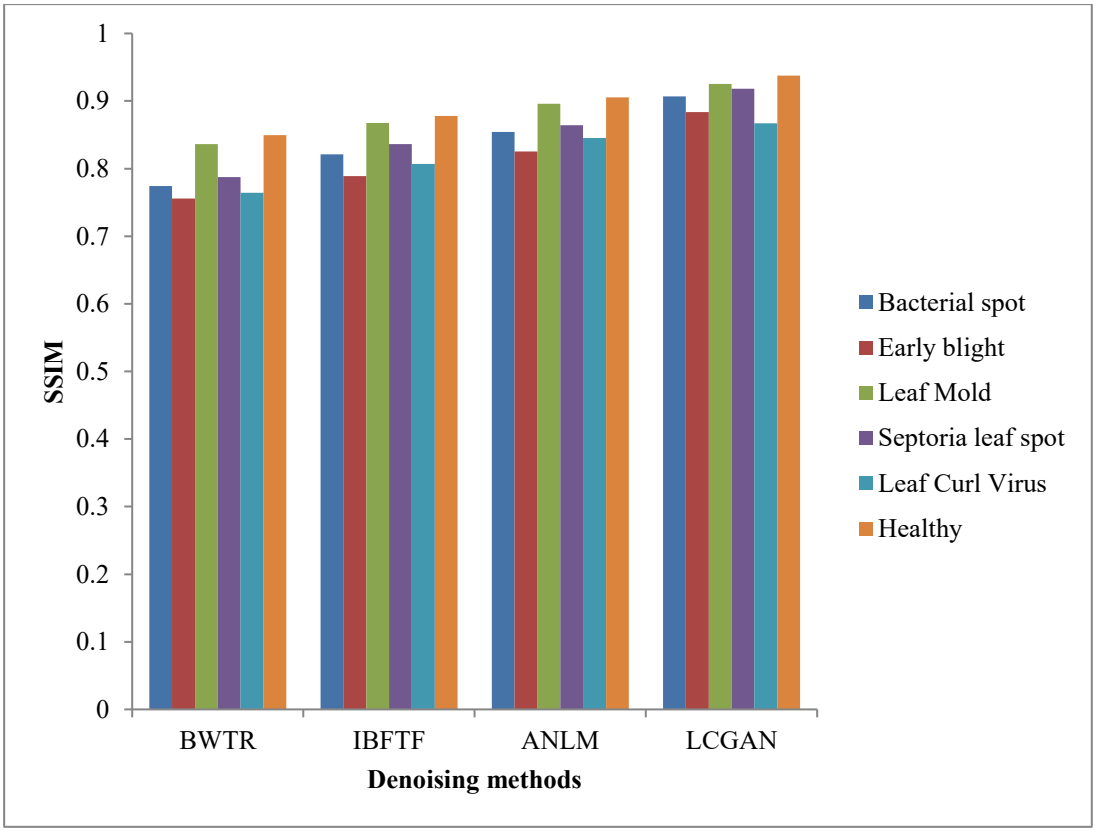

FIGURE 6. SSIM COMPARISON VS. DENOISING METHODS

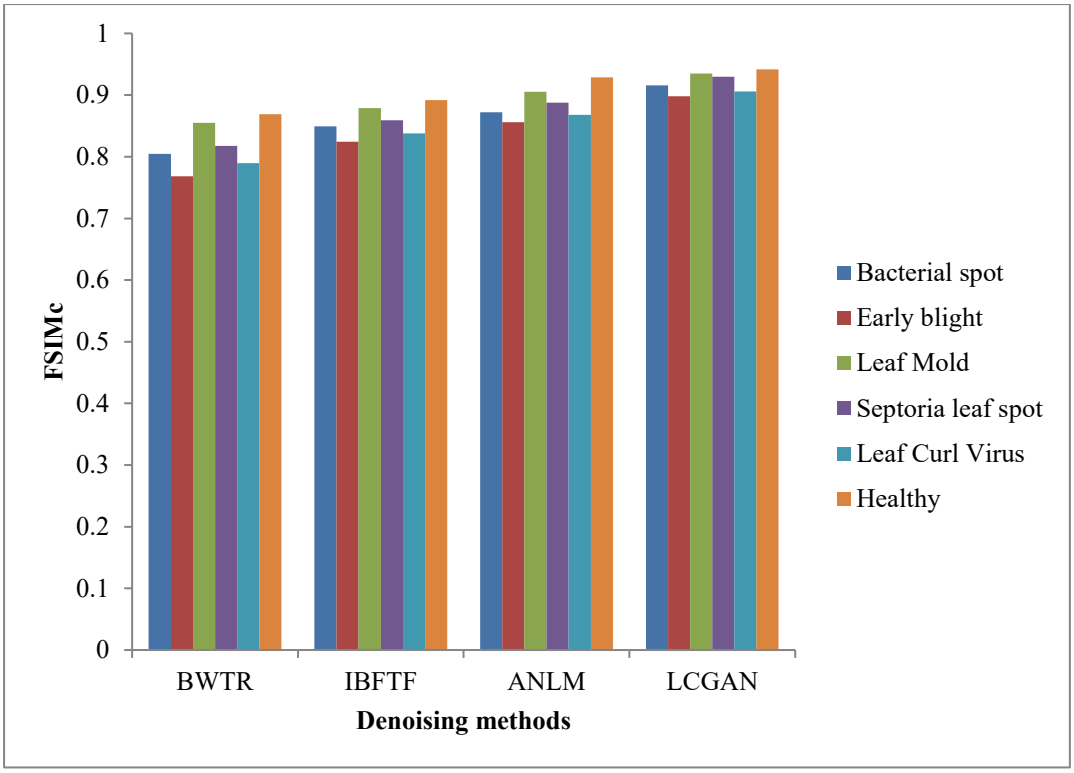

FIGURE 7. FSIMc COMPARISON VS. DENOISING METHODS

SSIM comparison of denoising methods such as BWTR, IBFTF, ANLM, and LCGAN among six classes is illustrated in Figure 6. LCGAN gives the highest SSIM of 0.9067, 0.8835, 0.9254, 0.9183, 0.8673, and 0.9378 among bacterial spot, early blight, leaf mold, Septoria leaf spot, leaf curl virus, and healthy. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.8497, 0.8781, and 0.9056 for healthy. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.7741, 0.8212, and 0.8545 for bacterial spots. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.7557, 0.7891, and 0.8254 for early blight. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.8361, 0.8675, and 0.8962 for leaf mold. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.7877, 0.8364, and 0.8641 for septoria leaf spot. BWTR, IBFTF, and ANLM give the lowest SSIM of 0.7641, 0.8069, and 0.8452 for the leaf curl virus. An FSIMc comparison of denoising methods such as BWTR, IBFTF, ANLM, and LCGAN among six classes is illustrated in figure 7. LCGAN gives the highest FSIMc of 0.9158, 0.8983, 0.9348, 0.9296, 0.9057, and 0.9418 among bacterial spot, early blight, leaf mold, Septoria leaf spot, leaf curl virus, and healthy. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.8692, 0.8918, and 0.9289 for healthy. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.8044, 0.8491, and 0.8719 for bacterial spot. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.7681, 0.8244, and 0.8558 for early blight. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.8551, 0.8789, and 0.9054 for leaf mold. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.8175, 0.8589, and 0.8878 for septoria leaf spot. BWTR, IBFTF, and ANLM give the lowest FSIMc of 0.7898, 0.8376, and 0.8682 for the leaf curl virus.

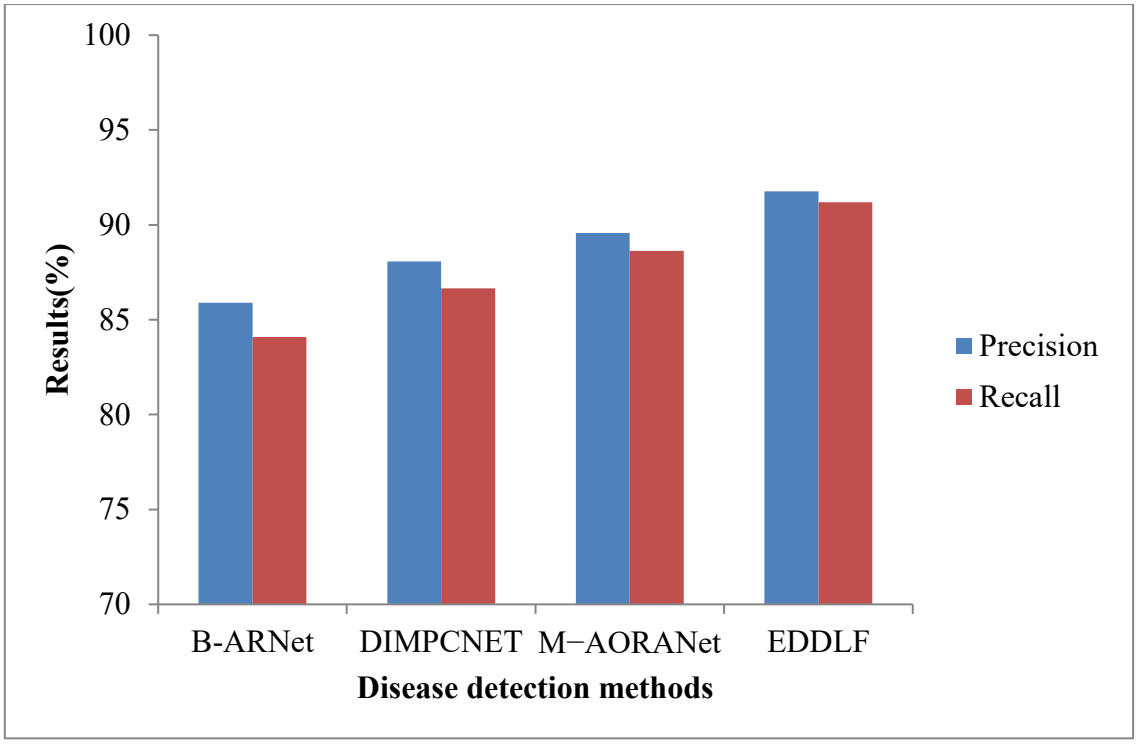

FIGURE 8. PRECISION AND RECALL COMPARISON VS. DETECTION METHODS

B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 among precision and recall metrics are illustrated in figure 8. B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 provide a precision of 85.89%, 88.08%, 89.57%, and 91.76%. B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 provide a recall of 84.09%, 86.65%, 88.63%, and 91.19%. B-ARNet, DIMPCNET, and M−AORANet have 5.86%, 3.68%, and 2.19% increased precision results when compared to MSDAIV3. B-ARNet, DIMPCNET, and M−AORANet have 7.10%, 4.54%, and 2.56% increased recall results when compared to MSDAIV3.

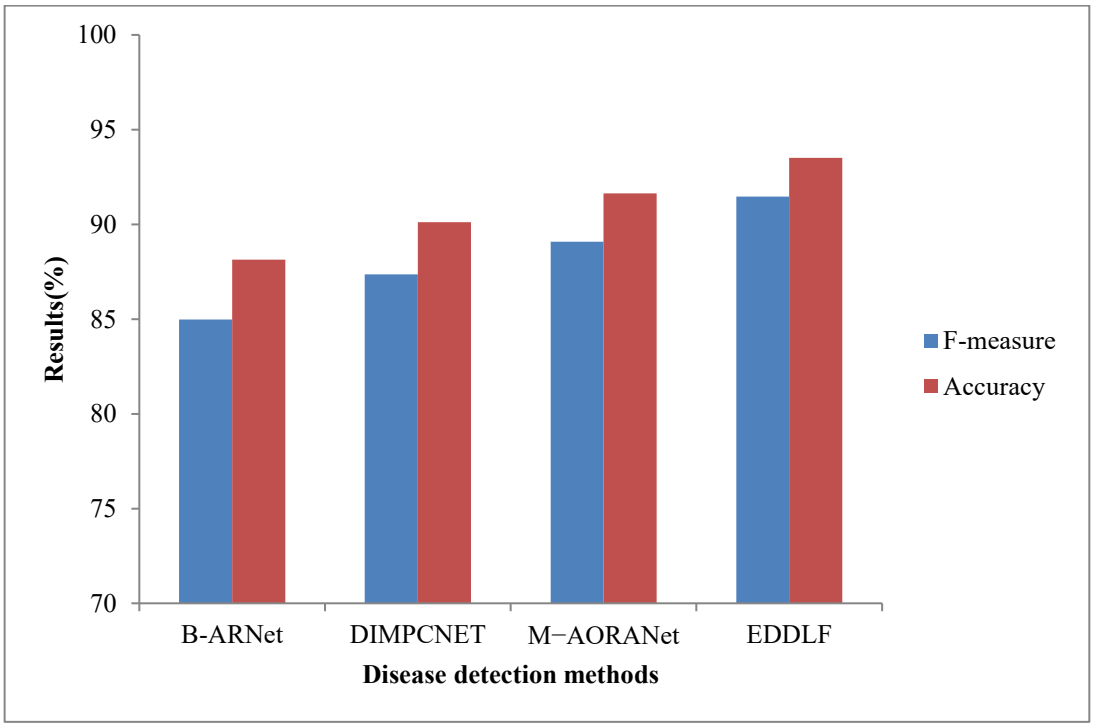

FIGURE 9. F-MEASURE AND ACCURACY COMPARISON VS. DETECTION METHODS

B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 among f-measure, and accuracy metrics are illustrated in figure 9. B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 provide an f-measure of 84.98%,

87.36%, 89.09%, and 91.47%. B-ARNet, DIMPCNET, M−AORANet, and MSDAIV3 provide an accuracy of 88.14%, 90.12%, 91.63%, and 93.52%. B-ARNet, DIMPCNET, M−AORANet has 6.49%, 4.11%, and 2.38% increased f-measure results when compared to MSDAIV3. B-ARNet, DIMPCNET, M−AORANet has 5.38%, 3.40%, and 1.89% increased accuracy results when compared to MSDAIV3.

5. CONCLUSION AND FUTURE WORK

In this paper, tomato leaf dataset is collected from the real-time dataset with six classes. Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced by enforcing an additional constraint and image is denoised from its actual image. LCGAN, generator sub-network is constructed using the densely connected networks, whereas multi-scale discriminator is proposed to leverage both local and global information to image. Secondly, three-segment linear conversion is utilized to increase image contrast. Thirdly, InceptionV3 model generates rich features from images, capturing both local and global aspects; it is enhanced by modified multi-scale spatial attention block with increased feature map. Finally, Multi-Scale Dual Attention Inception V3 (MSDAIV3) model is introduced with the integration of Channel Attention (CA) and Spatial Attention (SA). CA is introduced to utilize the inter-channel relationship of related features. MSA maps are generated by analyzing the inter-spatial relationship within the features. MSDAIV3 model is compared to current algorithms such as B-ARNet, DIMPCNET, and M−AORANet in terms of precision, recall, F-measure, and accuracy. Simulation shows that the proposed model is superior to other existing models in terms of automatically detecting tomato leaf disease. Denoising results of proposed model is compared to existing methods like BWTR, IBFTF, and ANLM using metrics such as PSNR, SSIM, and FSIMc. However, future research needs to further optimize the structure of model, computational efficiency, and validate its application effects in more practical scenarios.

REFERENCES

1. Trávníček, J., Schlatter, B., Helbing, M. and Willer, H., 2024. The World of Organic Agriculture 2025: Summary. In The World of Organic Agriculture. Statistics and Emerging Trends 2025 (pp. 20-31). Research Institute of Organic Agriculture FiBL and IFOAM−Organics International.
2. Savary, S.; Willocquet, L.; Pethybridge, S.J.; Esker, P.; McRoberts, N.; Nelson, A. The global burden of pathogens and pests on major food crops. Nat. Ecol. Evol. **2019**, 3, 430–439.
3. Panno, S., Davino, S., Caruso, A.G., Bertacca, S., Crnogorac, A., Mandić, A., Noris, E. and Matić, S., 2021. A review of the most common and economically important diseases that undermine the cultivation of tomato crop in the mediterranean basin. Agronomy, 11(11), pp.1-45.
4. Demilie, W.B., 2024. Plant disease detection and classification techniques: a comparative study of the performances. Journal of Big Data, 11(1), pp.1-43.
5. Ngugi, H.N., Akinyelu, A.A. and Ezugwu, A.E., 2024. Machine Learning and Deep Learning for Crop Disease Diagnosis: Performance Analysis and Review. Agronomy, 14(12), pp.1-39.
6. Sharma, S., Sharma, G. and Menghani, E., 2025. Tomato plant disease detection with pretrained CNNs: Review of performance assessment and visual presentation. Artificial Intelligence in Medicine and Healthcare, pp.67-85.
7. Abu John, M.; Bankole, I.; Ajayi-Moses, O.; Ijila, T.; Jeje, T.; Lalit, P.; Jeje, O. Relevance of Advanced Plant Disease Detection Techniques in Disease and Pest Management for Ensuring Food Security and Their Implication: A Review. Am. J. Plant Sci. 2023, 14, 1260–1295.
8. da Cunha, V.A.G., Hariharan, J., Ampatzidis, Y. and Roberts, P.D., 2023. Early detection of tomato bacterial spot disease in transplant tomato seedlings utilising remote sensing and artificial intelligence. Biosystems Engineering, 234, pp.172-186.
9. Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W. and Lin, C.W., 2020. Deep learning on image denoising: An overview. Neural Networks, 131, pp.251-275.
10. Zhang, K., Zuo, W. & Zhang, L. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Trans. Image Process. 27, 4608–4622 (2018).
11. Wang, C., Ren, C., He, X. & Qing, L. Deep recursive network for image denoising with global non-linear smoothness constraint prior. Neurocomputing 426, 147–161 (2021).
12. Sen, A. P. & Rout, N. K. A comparative analysis of the algorithms for de-noising images contaminated with impulse noise. Sens. Imaging 23, 11 (2022).
13. Chen, X., Zhou, G., Chen, A., Yi, J., Zhang, W. and Hu, Y., 2020. Identification of tomato leaf diseases based on combination of ABCK-BWTR and B-ARNet. Computers and Electronics in Agriculture, 178, p.105730.

14. Peng, D., Li, W., Zhao, H., Zhou, G. and Cai, C., 2023. Recognition of tomato leaf diseases based on DIMPCNET. Agronomy, 13(7), pp.1-24.
15. Islam, M.S., Sultana, S., Farid, F.A., Islam, M.N., Rashid, M., Bari, B.S., Hashim, N. and Husen, M.N., 2022. Multimodal hybrid deep learning approach to detect tomato leaf disease using attention based dilated convolution feature extractor with logistic regression classification. Sensors, 22(16), pp.1-31.
16. Zhou, C., Zhou, S., Xing, J. and Song, J., 2021. Tomato leaf disease identification by restructured deep residual dense network. IEEE Access, 9, pp.28822-28831.
17. Zhang, Y., Huang, S., Zhou, G., Hu, Y. and Li, L., 2023. Identification of tomato leaf diseases based on multi-channel automatic orientation recurrent attention network. Computers and Electronics in Agriculture, 205, p.107605.
18. Zhang, L., Zhou, G., Lu, C., Chen, A., Wang, Y., Li, L. and Cai, W., 2022. MMDGAN: A fusion data augmentation method for tomato-leaf disease identification. Applied Soft Computing, 123, p.108969.
19. Patil, M.A. and Manohar, M., 2022. Enhanced radial basis function neural network for tomato plant disease leaf image segmentation. Ecological Informatics, 70, p.101752.
20. Zhang, E., Zhang, N., Li, F. and Lv, C., 2024. A lightweight dual-attention network for tomato leaf disease identification. Frontiers in Plant Science, 15, pp.1-18.
21. Badiger, M. and Mathew, J.A., 2023. Tomato plant leaf disease segmentation and multiclass disease detection using hybrid optimization enabled deep learning. Journal of Biotechnology, 374, pp.101-113.
22. Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
23. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., 2017. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708.
24. Johnson, J., Alahi, A. and Fei-Fei, L., 2016. Perceptual losses for real-time style transfer and super-resolution. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 694-711). Springer International Publishing.