International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

WEIGHT HIPPOPOTAMUS OPTIMIZATION ALGORITHM MULTI THRESHOLD SEGMENTATION (WHOAMTS) AND MULTI-SCALE ATTENTION MULTI-AXIS VISION TRANSFORMER (MAMViT) 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

The rapid integration of artificial intelligence (AI) and computer vision into precision agriculture has revolutionized disease monitoring and crop management by enabling automated, accurate, and early detection of plant health conditions. In particular, tomato cultivation, which holds substantial economic importance worldwide, faces major productivity losses due to various foliar diseases. Early diagnosis of these diseases is crucial to prevent infection spread and minimize yield reduction. To tackle this challenge, this paper presents a comprehensive hybrid framework combining the Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS) and the Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT) for efficient tomato leaf disease detection. Initially, a Lagrange Conditional Generative Adversarial Network (LCGAN) is employed to pre-process and denoise tomato leaf images, producing high-quality data for subsequent segmentation. The LCGAN follows an encoder–decoder architecture guided by a Lagrangian constraint to minimize reconstruction error, effectively reducing illumination distortion and background noise. This enhances the visibility of diseased regions and simplifies the extraction of discriminative features for classification. After denoising, the WHOAMTS algorithm performs multi-threshold segmentation by determining optimal thresholds inspired by the natural behaviour of hippopotamuses— MAMViT such as defense mechanisms, territorial movement, and adaptive escape strategies. These behaviours are mathematically modelled to improve the global search balance and segmentation accuracy across multiple grayscale levels, ensuring precise region separation between infected and healthy tissues. Following segmentation, the Inception-V3 model is employed to extract abundant disease-specific features from the segmented regions, enabling rich feature representation for classification. The extracted features are then processed by the framework, which integrates convolutional feature encoding with multi-axis self-attention to capture both fine local structures and global dependencies. This dual-domain representation enables the network to recognize subtle visual cues in various

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

tomato leaf diseases, including Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, Leaf Curl Virus, and Healthy categories. Experimental analysis conducted on a real-time tomato leaf dataset confirms that the proposed LCGAN–WHOAMTS–Inception-V3–MAMViT pipeline achieves superior detection performance over conventional methods, with significant improvements in DSC, JSI, precision, recall, F-measure, and overall accuracy. The findings demonstrate the potential of the proposed framework as a robust and intelligent diagnostic system for sustainable and smart agricultural applications.

KEYWORDS: Lagrange Conditional Generative Adversarial Network (LCGAN), Weight Hippopotamus Optimization Algorithm (WHOA), Multi-Threshold Segmentation; Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT), Tomato Leaf Disease Detection, Deep Learning.

1. INTRODUCTION

Plant diseases pose a major threat to global agricultural productivity, food security, and economic stability. With the growing global population and the effects of climate change, early detection and precise management of plant diseases have become indispensable for achieving sustainable agricultural growth. Pathogenic infections can significantly reduce crop yield and quality if not detected at an early stage [1]. Traditional disease identification methods rely heavily on visual inspections by experts or laboratory-based chemical testing, which are often time-consuming, costly, and subjective. The urgent demand for real-time, scalable, and accurate detection methods has motivated researchers to leverage Artificial Intelligence (AI) and computer vision technologies to automate disease prediction and diagnosis. The integration of AI in agriculture has led to remarkable advancements in precision farming, where intelligent systems assist in disease forecasting, crop health monitoring, and decision-making support.

Over the past decade, plant disease prediction has increasingly relied on digital image analysis due to the widespread availability of high-resolution imaging devices [2,3]. By analysing leaf patterns, colour variations, and texture changes, machine learning algorithms can detect early symptoms invisible to the naked eye. Among different crops, tomatoes (Solanum lycopersicum) hold significant global importance because of their nutritional value and industrial applications. However, tomato plants are highly susceptible to various foliar diseases such as Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, and Leaf Curl Virus. These diseases usually manifest initially on the leaves, making leaf image analysis a reliable indicator of plant health. Hence, automating the detection of tomato leaf diseases using image-based deep learning models can substantially improve disease management, reduce pesticide usage, and ensure higher yield quality [4].

In the domain of computer vision for agriculture, image segmentation is a crucial pre-processing step for isolating diseased regions from healthy tissues. Effective segmentation allows subsequent classification models to focus only on relevant areas containing disease symptoms. However, agricultural images typically suffer from complex lighting conditions, occlusions, and non-uniform backgrounds, making segmentation [5] a challenging task. Traditional thresholding-based segmentation methods, such as Otsu's algorithm or fuzzy C-

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

means clustering, often fail when dealing with multiple intensity distributions and irregular boundaries. To overcome these limitations, multi-threshold segmentation techniques have been developed to divide the image into several meaningful grayscale levels. The success of multi-threshold segmentation largely depends on the ability to determine optimal threshold values that separate distinct regions accurately [6]. However, since the threshold search space grows exponentially with the number of gray levels, traditional methods may converge slowly or yield suboptimal results.

To improve segmentation accuracy, metaheuristic optimization algorithms inspired by biological and physical processes have been applied to automatically determine the optimal threshold combinations. Algorithms such as Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and Grey Wolf Optimizer (GWO) have shown promising results in image segmentation tasks. Building upon these strategies, this work introduces the Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS), which draws inspiration from the behavioural dynamics of hippopotamuses in nature. Hippopotamuses exhibit intelligent and cooperative behaviours such as group Défense, territory protection, and adaptive movement in water. These behaviours are mathematically modelled to balance exploration and exploitation during the optimization process, allowing WHOAMTS to effectively find optimal threshold values for precise segmentation. As a result, diseased and healthy leaf regions are distinctly separated across multiple gray levels, providing a strong foundation for downstream feature extraction and classification.

Although accurate segmentation helps isolate infected regions, deep and discriminative feature extraction is essential to differentiate between various disease categories [7]. For this purpose, Inception-V3, a deep convolutional neural network, is employed after segmentation to extract multi-scale spatial and texture features from the segmented leaf regions. The Inception-V3 architecture utilizes parallel convolutional kernels of different sizes within each layer, enabling it to capture local texture patterns as well as broader lesion structures. This multi-scale feature extraction property is particularly beneficial for plant disease analysis, where disease symptoms vary in size, shape, and colour intensity. The extracted features serve as a high-dimensional representation of the disease characteristics, allowing for improved classification performance when processed by advanced attention-based models [8].

Despite the success of convolutional neural networks (CNNs) like Inception-V3, they inherently suffer from limitations in modelling long-range dependencies due to their localized receptive fields. CNNs primarily focus on spatially adjacent features and thus struggle to capture global contextual relationships that may be crucial in understanding disease spread across an entire leaf surface. To address these limitations, Vision Transformers (ViTs) have emerged as a powerful alternative in computer vision. ViTs utilize self-attention mechanisms to compute relationships among all image regions, thereby learning both local and global dependencies. However, standard transformers are computationally expensive because the self-attention operation scales quadratically with image size, making them inefficient for high-resolution agricultural imagery where disease details appear at multiple scales.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

To overcome these challenges, this research proposes the Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT), an advanced vision transformer framework that efficiently integrates multi-scale feature learning with multi-axis attention mechanisms. The MAMViT model incorporates both block attention (for localized detail learning) and grid attention (for global feature interaction) within the same architecture. This dual-axis attention mechanism enables MAMViT to simultaneously capture fine lesion textures and broader structural patterns across the entire leaf image. Furthermore, a multi-scale attention fusion module aggregates hierarchical features extracted at different depths, preserving spatial precision while enhancing contextual understanding. This hierarchical attention design allows MAMViT to outperform traditional CNNs and single-scale transformers in handling the complex visual characteristics of tomato leaf diseases.

Before segmentation and feature extraction, agricultural leaf images often contain noise, illumination variations, and background clutter that can mislead segmentation algorithms and degrade classification accuracy. To address this, a Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced as a pre-processing step in the proposed pipeline. The LCGAN leverages an encoder–decoder architecture and incorporates Lagrangian constraints to minimize reconstruction loss and maintain image consistency. This network effectively removes environmental noise, corrects lighting irregularities, and enhances the visual quality of diseased regions, ensuring that subsequent segmentation and attention modules operate on clean, informative data. As a result, the LCGAN improves the robustness of the entire detection framework under real-world imaging conditions.

In summary, the proposed LCGAN–WHOAMTS–Inception-V3–MAMViT pipeline provides a unified, multi-stage approach for robust tomato leaf disease detection. The LCGAN ensures clean, noise-free inputs; the WHOAMTS performs optimal multi-threshold segmentation guided by bio-inspired optimization; Inception-V3 extracts abundant and discriminative disease features; and MAMViT integrates multi-scale and multi-axis attention mechanisms to classify diseases accurately. Experimental evaluations on real-time tomato leaf datasets demonstrate that the proposed hybrid model achieves superior performance compared to conventional deep learning methods in terms of accuracy, precision, recall, and F-measure. The combination of optimization-based segmentation, deep feature extraction, and attention-driven classification establishes a powerful, generalizable framework for AI-driven agricultural diagnostics, contributing to sustainable and intelligent crop disease management.

2. LITERATURE REVIEW

Wu et al. [2022] [9] proposed the Disease Segmentation Detection Transformer (DS-DETR) to efficiently segment leaf disease spots with several improvements over DETR. Additionally, a damage assessment is performed based on the area ratio of segmented leaves to disease spots. Initially, an unsupervised pre-training method was applied to DETR using the Plant Disease Classification Dataset (PDCD) to address the long training epochs and slow convergence of DETR. This pre-training allows the Transformer to learn leaf disease features in advance, and loading the pre-trained weights into DS-DETR accelerates convergence. Subsequently,

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Spatially Modulated Co-Attention (SMCA) was employed to assign Gaussian-like spatial weights to the query boxes, enabling different positions in the image to be trained with different importance levels, thereby improving model accuracy. Finally, an improved relative position encoding was incorporated into the Transformer structure to enhance spatial location representation. The DS-DETR model was evaluated on the Tomato Leaf Disease Segmentation Dataset (TDSD) using metrics including APmask, mean Average Precision (mAP), Intersection over Union (IoU), Precision, Recall, and F1-score.

Karthik et al. [2020] [10] presented two deep learning architectures for identifying types of infections in tomato leaves using the PlantVillage Dataset. The first architecture leveraged residual learning to extract significant features, while the second applied an attention mechanism on top of the residual network. The evaluation was performed using accuracy, Precision, Recall, and F1-score, demonstrating high classification performance across all disease categories.

hmad et al. [2020] [11] investigated tomato leaf disease classification using VGG-16, VGG-19, ResNet, and Inception V3, evaluated on both laboratory-based and field-collected datasets. The models were evaluated using accuracy, precision, recall, and F1-score, highlighting differences in performance between controlled and real-world environments, with Inception V3 identified as the most effective.

Ni et al. [2023] [12] introduced TomatoNet, an improved CNN based on ResNet18 enhanced with a squeeze-and-excitation (SE) module and a modified classifier structure, evaluated using the PlantVillage Dataset. The model was assessed with accuracy, Precision, Recall, F1-score, and Confusion Matrix, showing improved recognition performance over original ResNet18 and AlexNet networks.

Qi et al. [2022] [13] proposed an improved SE-YOLOv5 network for recognizing tomato virus diseases using greenhouse images. By incorporating a squeeze-and-excitation (SE) module, the network effectively extracted key features inspired by human visual attention. The model was evaluated using mean Average Precision (mAP), Precision, Recall, and F1-score, demonstrating improved detection performance compared to Faster R-CNN, SSD, and original YOLOv5.

Wang et al. [2024] [14] developed a tomato leaf disease detection method based on attention mechanisms and multi-scale feature fusion, evaluated on the PlantDoc Dataset and a tomato leaf disease dataset. The method integrates Convolutional Block Attention Module (CBAM) and BiRepGFPN in YOLOv6. Evaluation metrics included Precision, Recall, F1-score, and mean Average Precision (mAP), demonstrating improved detection performance, particularly for small lesion features.

Sunil and Jaidhar [2023] [15] proposed a Multilevel Feature Fusion Network (MFFN) combining ResNet50 with adaptive attention mechanisms (channel, spatial, and pixel attention). The model was evaluated on a tomato plant leaves dataset using accuracy, Precision,

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Recall, and F1-score, showing superior performance compared to existing approaches. Additionally, a pesticide recommendation module was proposed to provide guidance based on the classified disease type.

Chowdhury et al. [2021] [16] proposed a deep learning framework based on the EfficientNet architecture for tomato disease classification using 18,161 plain and segmented tomato leaf images from the PlantVillage (PV) Dataset. Two segmentation models—U-Net and Modified U-Net—were employed for leaf segmentation prior to classification. The study reported comparative analyses for three classification scenarios: binary classification (healthy vs. unhealthy leaves), six-class classification (healthy and grouped diseased leaves), and ten-class classification (healthy and individual diseased leaf types). The Modified U-Net model was evaluated using Accuracy, Intersection over Union (IoU), and Dice Coefficient for segmentation performance. For classification, various versions of EfficientNet were trained, where EfficientNet-B7 achieved the best results for binary and six-class classification, and EfficientNet-B4 performed best for ten-class classification. The study concluded that deeper network architectures trained on segmented images significantly improved disease classification performance compared to unsegmented images.

3. PROPOSED METHODOLOGY

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 12s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

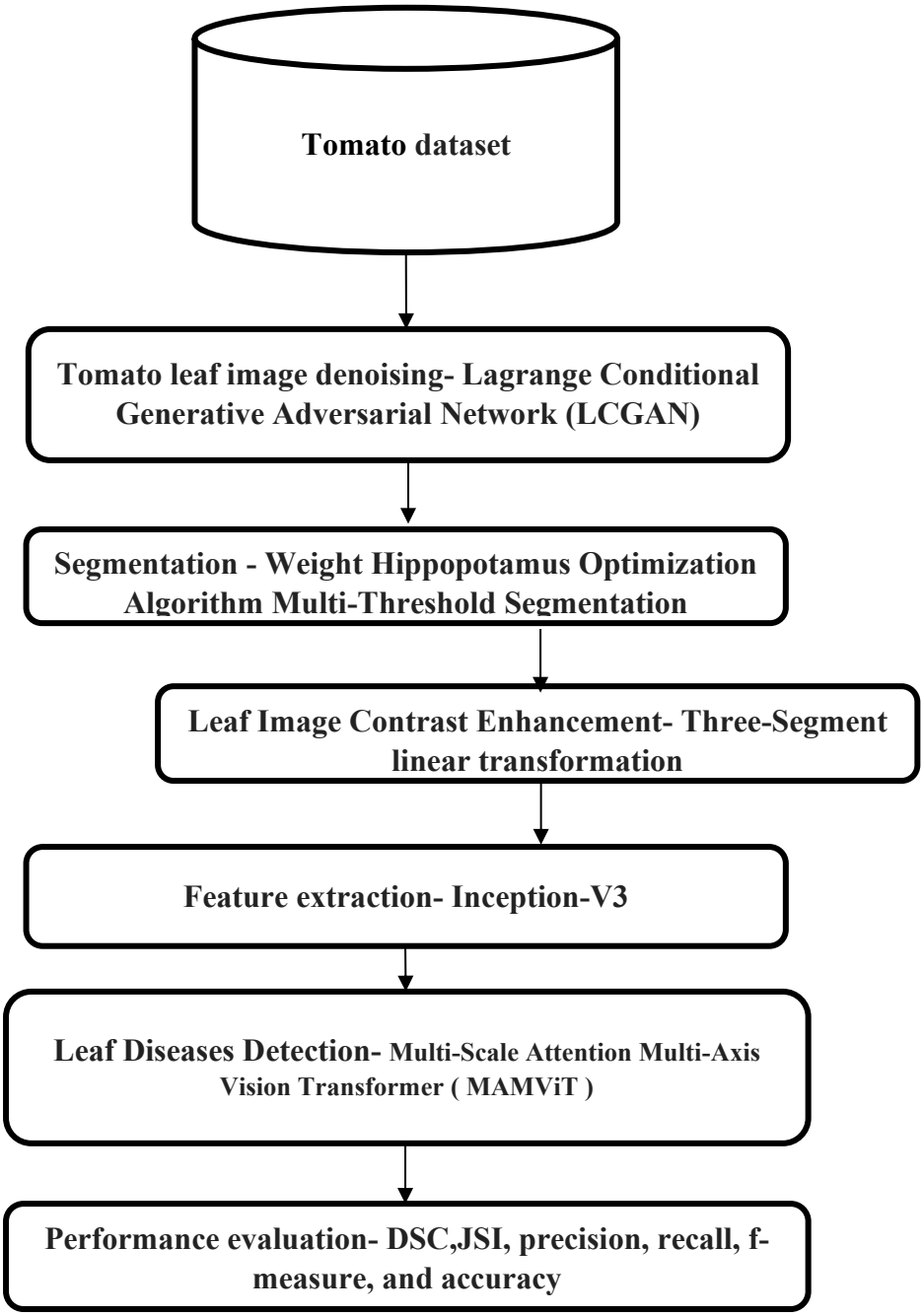


FIGURE 1. PROPOSED TOMATO LEAF DISEASES DETECTION MODEL

The proposed methodology introduces an advanced tomato leaf disease detection framework integrating multiple intelligent components for optimal performance. Initially, Lagrange Conditional Generative Adversarial Network (LCGAN) is employed for image denoising to enhance visual quality. The Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS) effectively segments the diseased leaf regions. A three-segment linear transformation improves image contrast, followed by Inception-V3 for rich feature extraction (figure 1). Finally, the Multi-Scale Attention Multi-Axis Vision Transformer

Received: August 08 2025

International Journal of Applied Mathematics

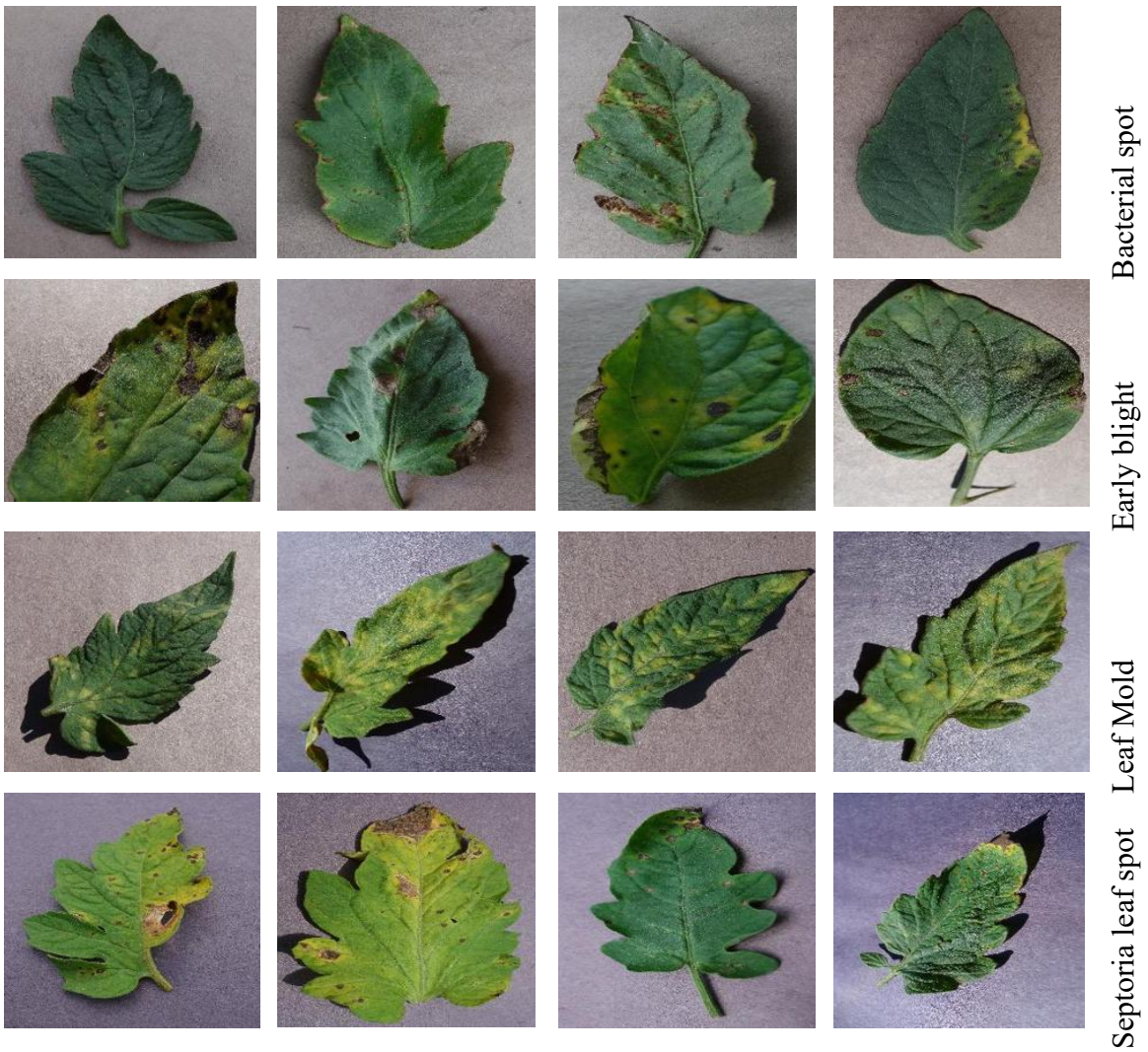
Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(MAMViT) accurately classifies leaf diseases, with performance evaluated using DSC,JSI, precision, recall, F-measure, and accuracy.

3.1. DATASET COLLECTION

In this study, the tomato leaf dataset was gathered from real-time sources and used for image processing and computer vision applications. The public dataset originally includes nine categories representing both healthy and diseased tomato leaves. For this research, six major classes were selected—Bacterial spot, Early blight, Leaf Mold, Septoria leaf spot, Leaf Curl Virus, and Healthy leaves—as illustrated in Figure 2. Each image follows the Red-Green-Blue (RGB) color model with an original resolution of 256×256 pixels, which was resized to 224×224 pixels for uniform processing. Backgrounds were removed prior to model training to enhance feature clarity [17] . In total, 10,239 tomato leaf images were utilized, comprising 1,702 Bacterial spot, 800 Early blight, 762 Leaf Mold, 1,417 Septoria leaf spot, 4,286 Leaf Curl Virus, and 1,272 Healthy leaf samples, all stored in JPG format.



Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

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

As shown in Table 1, the dataset was organized into three subsets: 70% for training, 20% for testing, and 10% for validation. Tomato leaf images were captured in natural environmental conditions using a camera, after which an image enhancement algorithm was applied to remove noise and improve the visual quality of the original images.

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

The Lagrange Conditional Generative Adversarial Network (LCGAN) is introduced to effectively remove noise from tomato leaf images while maintaining fine structural details. An additional Lagrange constraint ensures that the denoised output closely matches the ground truth image, while the adversarial loss improves overall perceptual quality [18]. The framework includes three main components—generator, multi-scale discriminator, and refined perceptual loss function—to achieve high-quality noise removal.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

LCGAN Architecture

The generator G is designed as a densely connected symmetric deep CNN with skip connections, enabling efficient feature propagation and better convergence. It learns an end-to-end mapping from noisy to clean images. The discriminator D is a multi-scale convolutional network that differentiates between real (clean) and generated (denoised) images using hierarchical contextual features. The combined adversarial training ensures the generation of realistic and visually enhanced outputs.

Generator Design

The generator employs dense blocks and transition layers (down-sampling, up-sampling, and non-sampling). Skip connections across symmetric layers enhance gradient flow and preserve image details [19].

TABLE 2. SIMPLIFIED GENERATOR ARCHITECTURE

Layer	Description	Channels (nc)
Input	RGB Image	3
Conv3×3 + BN + ReLU + MaxPool	Feature Extraction	64
Dense Block (4) + Td	Down-sampling	128
Dense Block (6) + Td	Down-sampling	256
Dense Block (8) + Tn	No-sampling	512
Dense Block (6) + Tu	Up-sampling	120
Dense Block (4) + Tu	Up-sampling	64
Conv3×3 + Tanh	Output Reconstruction	3

Multi-Scale Discriminator

The discriminator network is composed of stacked convolutional layers with Batch Normalization and PReLU activations, followed by a multi-scale pooling module. This structure incorporates both local and global contextual cues to determine whether the denoised image is authentic or synthesized.

TABLE 3. SIMPLIFIED DISCRIMINATOR ARCHITECTURE

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Layer	Description	Channels (nc)
Input	Noisy/Clean Image Pair	6
Conv3×3 + BN + ReLU + MaxPool	Feature Extraction	64–512
Multi-Scale Pooling (4 Levels)	Global Context	72
Sigmoid Activation	Classification Output	1

Refined Perceptual Loss

To enhance both quantitative and visual quality, a refined loss function integrates pixel-level, perceptual, and adversarial losses:

$$L_{RP} = L_E * \lambda + \lambda_A L_A + \lambda_P L_P \quad (1)$$

Here L_E is the Euclidean loss between denoised and ground truth images, L_P is the perceptual loss ensuring feature-level similarity, and L_A represents the adversarial feedback from the discriminator (Equation 1). The Lagrange multipliers λ_P, λ_A balance these components to produce smooth and visually consistent denoised images. Once denoised, the images are further processed through contrast enhancement to improve visual detail before segmentation.

3.3. THREE-SEGMENT LINEAR TRANSFORMATION BASED CONTRAST ENHANCEMENT

Three-segment linear conversion is utilized to increase image contrast. Equation (2) is used for the three-segment linear transformation [20],

$$f(i,j) = \begin{cases} k_1 \times d(i,j), & 0 \le d(i,j) < a \\ b + k_2 \times d(i,j), & a \le d(i,j) \le c \\ d + k_3 \times d(i,j), & c < d(i,j) \le 255 \end{cases} \quad (2)$$

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,

$$k_1 = \frac{a}{b}, k_2 = \frac{d-b}{c-a}, k_3 = \frac{255-d}{255-c} \quad (3)$$

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(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. WEIGHT HIPPOPOTAMUS OPTIMIZATION ALGORITHM MULTI-THRESHOLD SEGMENTATION (WHOAMTS)

The **Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS)** is a bio-inspired optimization strategy developed to achieve accurate and efficient segmentation of tomato leaf disease images [21]. The algorithm imitates the natural behaviors of hippopotamuses, such as exploration, defense, and escape, to balance global and local search capabilities. By incorporating an **Adaptive Weight Mechanism (AWM)**, WHOAMTS dynamically controls exploration and exploitation intensity during the optimization process. This adaptive weighting improves convergence stability and segmentation accuracy, especially in complex tomato leaf images with varied color tones and texture irregularities.

Initialization

The segmentation process begins with initializing a population of hippopotamuses, where each hippopotamus represents a candidate solution consisting of multiple threshold values. These thresholds are distributed randomly within the search space bounded by the image's intensity range to ensure diversity at the start of optimization.

$$x_{ij} = lb_j + r \times \left(ub_j - lb_j\right) \quad (4)$$

Here, x_{ij} denotes the position of the ith hippopotamus in the jth dimension, r is a random number in [0,1], and ub_j, lb_j represent the upper and lower bounds of the search space, respectively (Equation 4). In the tomato leaf disease dataset, these bounds correspond to pixel intensity limits within RGB or grayscale channels.

Exploration Phase

In the exploration stage, hippopotamuses move across wide regions of the search space in search of optimal food sources (potential thresholds). The process is influenced by the dominant hippopotamus (the best-performing solution) but also integrates randomness for diversity preservation. The adaptive weight factor y_1 determines the step size and direction of movement, allowing the algorithm to dynamically adjust exploration intensity based on iteration progress.

$$x_{ij}(t+1) = x_{ij}(t) + y_1 \times \left(D_{\text{hippo}} - I_1 \times x_{ij}(t)\right) \quad (5)$$

Where D_{hippo} represents the dominant hippopotamus (global best solution), I_1 is the dominance factor, and y_1 is an adaptive weight that controls the exploration rate (Equation 5). This phase

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

prevents premature convergence and ensures the algorithm explores various potential threshold sets for precise segmentation of diseased regions in tomato leaves.

Adaptive Weight Mechanism (AWM)

The Adaptive Weight Mechanism plays a central role in maintaining the balance between exploration and exploitation. Initially, a higher weight value is assigned to encourage broad exploration of the search space [22]. As iterations progress, the weight gradually decreases, shifting the search toward local exploitation and fine-tuning. The adaptive weight wt at iteration t is computed as:

$$\mathrm{wt} = \omega_{max;} - \frac{(\omega_{max;-}\omega_{min})\times\ t}{t_{max}} \quad (6)$$

Here, $\omega_{max;-}\omega_{min}$ represent the maximum and minimum weight values, respectively, and t_{max} denotes the total number of iterations (Equation 6). This dynamic adjustment ensures that early iterations focus on global search, while later stages emphasize refinement near optimal solutions. In tomato leaf segmentation, AWM enables accurate threshold determination under diverse lighting and texture conditions, improving disease spot detection accuracy.

Predator Defense Phase

When threatened, hippopotamuses exhibit strong defensive behavior, such as vocalizations or movement toward safer zones. In WHOAMTS, this phase enhances the algorithm's diversification ability by relocating some candidate solutions toward unexplored search regions. This prevents stagnation and promotes robustness in high-dimensional segmentation problems.

$$\mathrm{Predator}_j = lb_j + r \times \left(ub_j - lb_j\right) \quad (7)$$

This mechanism ensures the continual introduction of new potential threshold candidates, helping the algorithm identify meaningful pixel intensity boundaries that distinguish infected leaf areas from healthy ones.

Escape Phase (Exploitation)

If the defense mechanism is insufficient, hippopotamuses escape to safer water regions, representing the **exploitation** phase of the algorithm. This phase focuses on refining solutions around the best-performing regions in the search space. Using adaptive weights, the step size is controlled to ensure fine-grained updates near optimal threshold values.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$x_{ij}(t+1) = x_{ij}(t) + r \times \left(lb_j^{\text{local}} + s_1 \times \left| \left(ub_j^{\text{local}} - lb_j^{\text{local}} \right) \right| \right) \quad (8)$$

Here, ub_j^{local} and lb_j^{local} define the local search boundaries (Equation 8) , while s_1 is the step-size control parameter. This step enhances segmentation precision by capturing subtle disease characteristics such as mild discoloration or small lesion regions within tomato leaves.

Fitness Evaluation and Optimal Segmentation

Each candidate solution (set of thresholds) is evaluated using a fitness function—commonly measure segmentation quality [23]. The weighted optimization process iteratively updates hippopotamus positions until the best threshold values are identified. These optimal thresholds are then applied to the input image, segmenting it into multiple regions representing segmentation (Figure 3). The result is a high-quality segmented image that enables accurate feature extraction for subsequent classification by the Inception-V3 and MAMViT models.

Algorithm: Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS)

Step 1: Input the tomato leaf image I and convert it to grayscale if necessary.
Step 2: Initialize algorithm parameters including population size (N), number of thresholds (L), maximum iterations (t_max), and weight limits (w_max, w_min).
Step 3: Define lower (lb) and upper (ub) intensity bounds based on image pixel range.
Step 4: Initialize the population of hippopotamuses (candidate threshold sets) randomly.
Step 5: Evaluate the fitness of each hippopotamus using a segmentation quality metric (such as Otsu's variance).
Step 6: For each iteration t = 1 to t_max, compute adaptive weight.
Step 7: Identify the dominant hippopotamus (best fitness).
Step 8: Perform exploration phase.
Step 9: Execute predator defense phase: introduce new random individuals using Predator.
Step 10: Conduct escape phase (exploitation).
Step 11: Recalculate fitness for all hippopotamuses and update the dominant hippopotamus if a better one is found.
Step 12: Repeat until the maximum number of iterations (t_max) is reached.
Step 13: Output the optimal threshold values T_opt = D_hippo.
Step 14: Apply T_opt to the image for multi-threshold segmentation to obtain segmented regions.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

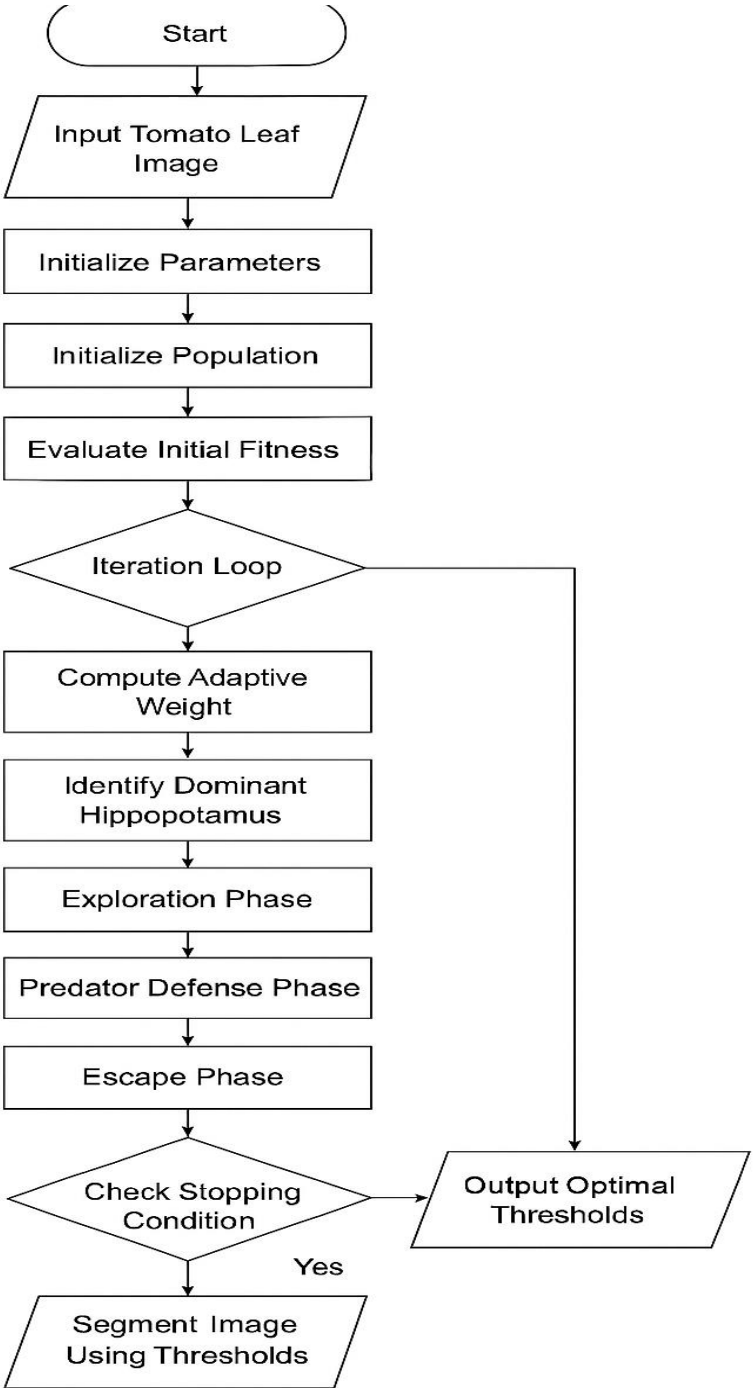


FIGURE 3. FLOWCHART OF WHOA

3.5. FEATURE EXTRACTION BY INCEPTION V3

After the segmented tomato leaf images are obtained from the WHOAMTS module, the Inception-V3 network is employed to extract deep discriminative features that characterize disease patterns. Each segmented image, resized to 224 × 224 × 3, is fed into the pretrained

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Inception-V3 model, which uses parallel convolutional kernels of varying receptive fields (1×1, 3×3, 5×5) to capture both fine-grained lesion textures and global structural information. The early convolutional layers of Inception-V3 focus on low-level edge and color features, while deeper layers encode high-level semantic patterns associated with different disease symptoms such as Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, and Leaf Curl Virus. By employing factorized convolutions, batch normalization, and auxiliary classifiers, Inception-V3 reduces computational complexity while preventing overfitting [24]. The output feature maps are taken from the global average pooling (GAP) layer to produce a compact feature vector for each leaf image. These vectors serve as the input for the MAMViT classifier, which performs multi-scale attention-based disease recognition

3.6. LEAF DISAESE DETECTION BY MULTI-SCALE ATTENTION MULTI-AXIS VISION TRANSFORMER (MAMVIT)

The Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT) framework represents a modern and comprehensive paradigm in visual feature learning, integrating both local feature discrimination and global dependency modeling within a unified structure [24]. In conventional convolutional neural networks (CNNs), the receptive field is spatially restricted by the size of the convolutional kernel. Although stacking multiple convolutional layers can expand this field, such networks often struggle to model long-range dependencies that are essential for understanding global relationships within images. On the other hand, Vision Transformers (ViTs) employ self-attention mechanisms that excel at capturing wide-ranging contextual dependencies across the entire image. However, their purely attention-based structure lacks the inductive biases of CNNs—such as locality, translation equivariance, and hierarchical abstraction—that are critical for spatial precision.

To overcome these limitations, MAMViT introduces a hybridized transformer–convolution structure, where a multi-axis attention mechanism enables simultaneous local and global feature encoding. The block attention component focuses on fine-grained local features, while the grid attention component extends the model's awareness to distant regions within the feature space. Furthermore, a multi-scale attention fusion mechanism aggregates hierarchical representations to produce rich, semantically consistent feature maps that are crucial for dense prediction tasks such as segmentation and object recognition.

3.6.1. Overall Architecture

The overall architecture of the MAMViT framework follows an encoder–decoder configuration that enables both feature abstraction and spatial reconstruction. The encoder progressively compresses the input information through successive convolutional and transformer blocks, thereby capturing hierarchical representations from fine textures to high-level semantics. The decoder, conversely, expands the feature maps to reconstruct a detailed spatial output, such as a segmentation map. Initially, the input image $X \in R^{H \times Wx3}$ is divided into non-overlapping patches through a stem convolution layer, producing low-level feature

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

embeddings $F_0 \in R^{\frac{H}{2}X\frac{W}{2}XC}$.These patches act as tokens for transformer-based operations. The encoder then employs MBConv (mobile inverted bottleneck convolution) layers for efficient feature compression, followed by MaxViT blocks that perform multi-axis self-attention (Max-SA) operations [25]. Formally, the forward propagation through the network can be expressed as:

$$\hat{y} = D\big(\varepsilon(x)\big) \tag{9}$$

where ε represents the encoder operator, responsible for extracting hierarchical multi-scale features, and D denotes the decoder that reconstructs pixel-wise predictions from these encoded representations (Equation 9).

The decoder uses patch expanding layers interleaved with attention modules to recover fine spatial details, while skip connections ensure that critical spatial information from earlier layers is preserved. To refine the feature alignment between encoder and decoder, MAMViT uses a Multi-Scale Channel-Spatial Attention (MCSA) unit that adaptively calibrates inter-stage features.

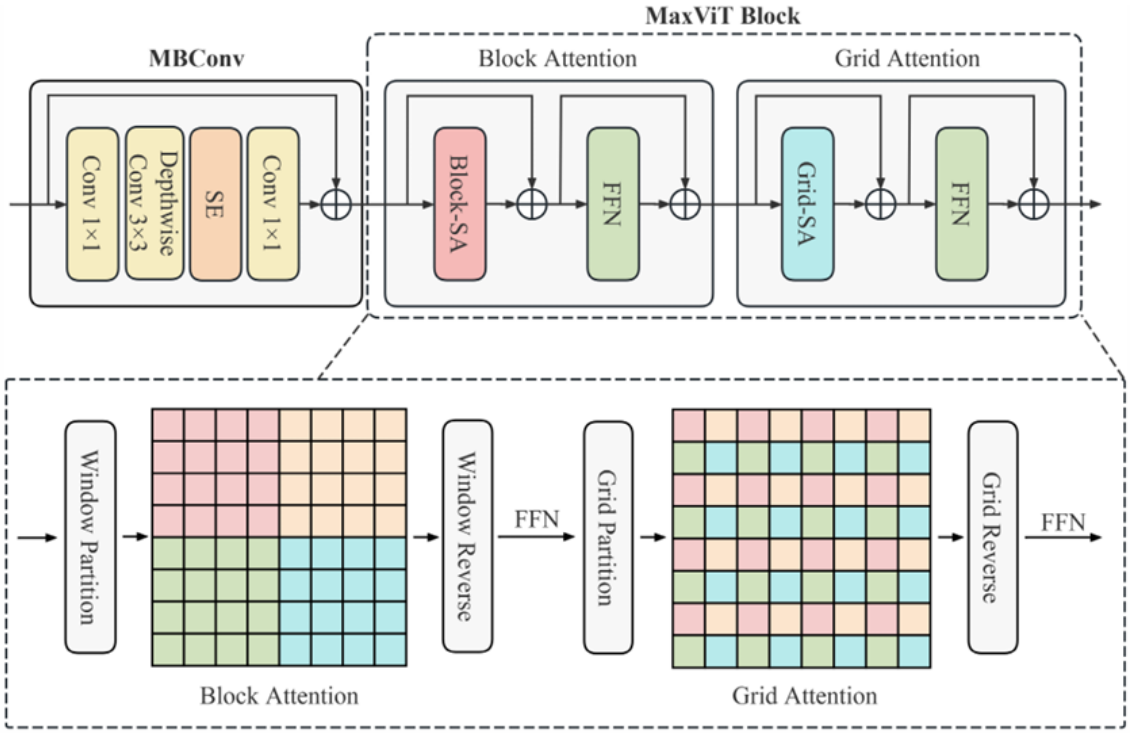


FIGURE 4. THE DIAGRAM ILLUSTRATES THE DETAILED STRUCTURES OF MBCONV AND MAXVIT BLOCK.

3.6.2. Multi-Axis Attention Mechanism

The **multi-axis attention mechanism** is the core innovation of MAMViT, designed to combine the strengths of both local and global feature processing within the same transformer block (figure 4). It alternates between two sequential attention operations: **block attention**, which

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

focuses on local receptive regions, and **grid attention**, which models distant interactions across the entire feature space [26].

3.6.2.1. Block Attention (Local Context Extraction)

Block attention partitions the feature map into small, non-overlapping windows of size $p \times p$ times. Each block independently computes self-attention to capture localized contextual relationships, allowing the network to focus on nearby textures, edges, and object boundaries. This operation can be expressed as:

$$\text{Block: } (H, W, C) \rightarrow \left(\frac{HW}{p^2}, p^2, C\right) \quad (10)$$

$$x \leftarrow x + Unblock(RelAttention(Block(LN(x)))) \quad (11)$$

where LN denotes layer normalization and "Unblock" restores the original spatial arrangement after attention computation in equation 11. The relative attention function, which integrates positional encoding, is defined as:

$$RelAttention(Q, K, V) = Softmax\left(\frac{Qk^T}{\sqrt{d}} + B\right)\text{V} \quad (12)$$

where Q, K, V are the query, key, and value matrices, d is the embedding dimension, and B is the learnable relative positional bias matrix as denoted in equation 12. This formulation enables the model to learn relationships not only based on feature similarity but also relative spatial distance.

3.6.2.2. Grid Attention (Global Context Modeling)

While block attention is efficient in capturing fine-grained local dependencies, it remains limited in handling large-scale contextual information. To complement this, **grid attention** rearranges the feature maps into a sparse, globally distributed grid of size G. Each grid cell attends to other non-adjacent cells, facilitating long-range feature correlation across distant spatial regions.

The transformation is given by:

$$\text{Grid: } (H, W, C) \rightarrow \left(G^2, \frac{HW}{G^2}, C\right) \quad (13)$$

and the corresponding attention computation follows:

$$x \leftarrow x + Ungrid(RelAttention(Grid(LN(x)))) \quad (14)$$

after which a feed-forward network (FFN) further refines the representation:

$$x \leftarrow x + MLP(LN(x)) \quad (15)$$

The combination of Equations defines a complete multi-axis transformer block, achieving both local and global dependency modelling within linear complexity.

3.6.3. Encoder and Feature Extraction

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The encoder is organized into multiple hierarchical stages, where each stage applies **MBConv** for spatial downsampling followed by a MaxViT block for attention-based feature enhancement. The MBConv unit expands, filters, and projects features efficiently:

$$x \leftarrow x + \text{Proj}(\text{SE}(\text{DWConv}(\text{Conv}(\text{Norm}(x)))))) \quad (16)$$

where *Conv* and *DWConv* represent 1×1 pointwise and 3×3 depthwise convolutions, respectively, and *SE* (Squeeze–Excitation) adaptively reweights channel responses (equation 16). When downsampling is required, the first MBConv of each stage uses stride-2 convolutions as:

$$x \leftarrow \text{Proj}(\text{Pool2D}(x)) + \text{Proj}(\text{SE}\left(\text{DWConv} \downarrow \left(\text{Conv}\left(\text{Norm}(x)\right)\right)\right) \quad (17)$$

This structure doubles channel capacity while halving spatial dimensions, allowing deeper feature abstraction without excessive parameter growth.

3.6.4. Multi-Scale Attention Fusion

To unify multi-level features extracted from different encoder stages, MAMViT incorporates a Multi-Scale Attention Fusion (MSA) module that aggregates spatially and semantically diverse information. [27] Let the features from different encoder stages be $\{F1,F2,F3\}$. The fused multi-scale representation is obtained by:

$$F_{MSA=} \sum_{i=1}^{3} \alpha i \, A(F_i), \quad (18)$$

where αi are adaptive weights obtained via softmax normalization, ensuring balanced contribution from each scale. A represents a two-stage channel–spatial attention operator in equation 18.

Channel Attention

Channel attention enhances discriminative channels while suppressing irrelevant ones. Given a feature tensor $F \in R^{C \times H \times W}$.

$$M_c(F) = \sigma(MLP(\text{AvgPool}(F) + \text{MLP}(\text{MaxPool}(F))), \quad (19)$$

where the shared multilayer perceptron (MLP) is parameterized. The sigmoid activation σ (·) normalizes attention weights between 0 and 1, and the refined output is given as:

$$F' = M_c(F) \odot F \quad (20)$$

This process adaptively scales feature channels based on global statistics, allowing the model to prioritize meaningful feature maps.

Spatial Attention

Spatial attention complements channel attention by identifying critical regions within the feature map that contribute most to the task objective. It is computed as in equation 21 & 22:

$$M_s(F') = \sigma(F^{7x7}([\text{AvgPool}(F'); \text{MaxPool}(F')])), \quad (21)$$

$$F'' = M_s(F') \odot F' \quad (22)$$

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

where $F^{7x7}(\cdot)$ represents a convolutional operation with a 7×7 kernel, and [;] indicates concatenation. The combined mechanism forms a comprehensive attention scheme that balances global semantics and local spatial precision. Figure 3, consists of two sequential sub modules: channel attention and spatial attention.

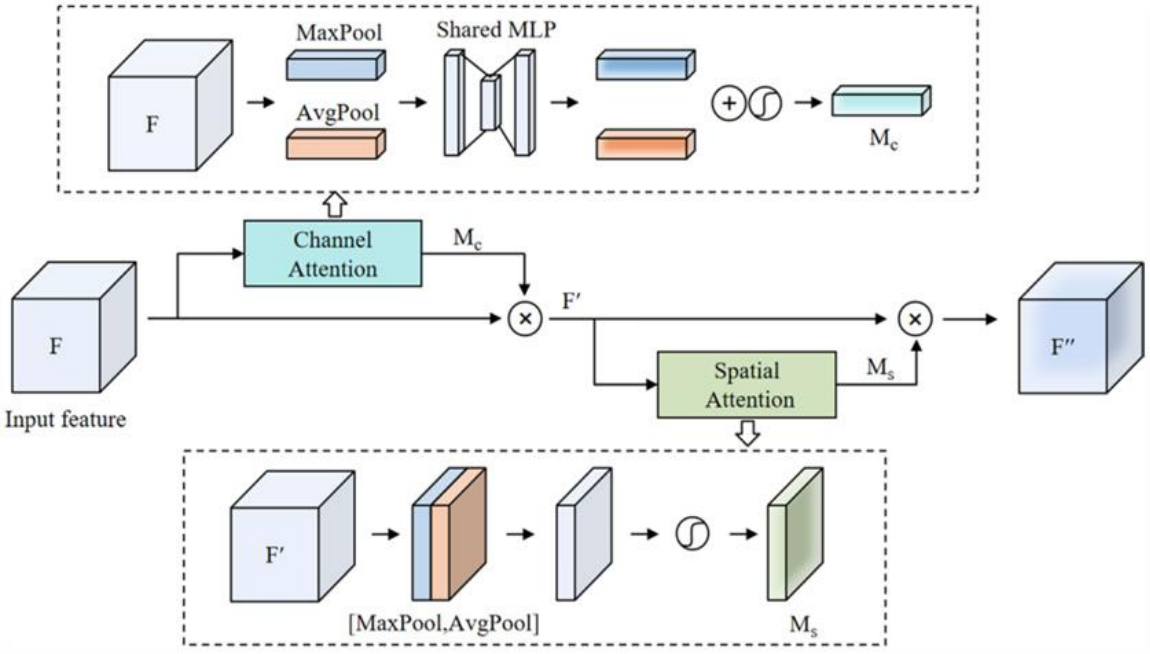


FIGURE 5.THE WORKFLOW OF CHANNEL AND SPATIAL ATTENTION ARCHITECTURE.

3.6.5. Multi-Scale Channel-Spatial Attention (MCSA) for Skip Connections

To preserve continuity between encoder and decoder layers while mitigating noise, MAMViT integrates **Multi-Scale Channel-Spatial Attention (MCSA)** within its skip connections [27,28]. This mechanism refines intermediate features by leveraging shallow, intermediate, and deep features simultaneously (figure 5).

Given shallow features F_1, current features F_0, and deep features F_2, the process is defined as in equation 23 to 25:

$$M_c(F_2) = \sigma(\text{Conv}(\text{AvgPool}(F_2)) + \text{Conv}(\text{MaxPool}(F_2))) \quad (23)$$

$$M_s(F_1) = \sigma\big(\text{Conv}(F_1)\big) \quad (24)$$

$$F_{refined} = (M_c(F_2) \odot F_0) \odot M_s(F_1) \quad (25)$$

This hierarchical attention reinforces the semantic depth of F_2 and structural clarity of F_1, resulting in a refined mid-level representation $F_{refined}$ that aligns feat3 ures across scales and stages.

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.6.6. Decoder Reconstruction and Output Generation

The decoder reverses the encoding process through **patch expanding layers** that perform learnable upsampling, restoring the spatial resolution while gradually integrating refined skip features. At each stage iii:

$$Y_{i-1} = PatchExpand(Y_i) + F_{refined}^{(i)} \quad (26)$$

The final decoding stage performs a 4× upsampling to produce an output that matches the original input dimensions H × W. The resulting feature maps are then passed through a softmax classifier to yield pixel-wise probability distributions. The Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT) effectively unifies the strengths of convolutional networks and transformers into a single model capable of multi-level contextual reasoning. Its dual-axis attention enables balanced perception between local structures and global semantics, while multi-scale fusion ensures information consistency across hierarchical layers. The MCSA-based skip refinement mitigates noise and enhances inter-layer feature alignment.

By maintaining linear computational complexity and preserving fine spatial detail, MAMViT delivers high segmentation accuracy across varied datasets and image modalities. This design paradigm can be extended beyond medical imaging to areas such as autonomous vision, satellite image analysis, and fine-grained object recognition, representing a robust step toward more adaptive and context-aware visual understanding frameworks.

3.7. EXPERIMENTATION AND RESULTS

This section presents a comprehensive performance comparison of the proposed Multi-scale Attention Multi-axis Vision Transformer (MAMViT) framework with existing detection and image denoising methods using the real-time tomato leaf dataset. The dataset consists of six categories of tomato leaves: Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, Leaf Curl Virus, and Healthy. The images were captured under natural environmental conditions using a high-resolution camera. To improve data quality, an image enhancement algorithm was applied to remove noise and enrich visual details, ensuring higher fidelity for model training. All experiments were carried out using MATLAB 2021a as the deep learning framework. The input image size was set to 224 × 224 pixels, and the neural network training process was accelerated using GPU computation.

3.8. EVALUATION INDEX

The disease detection results of the proposed MAMViT framework are compared with state-of-the-art methods using multiple quantitative and qualitative evaluation metrics. Segmentation performance is evaluated using dice similarity coefficient (DSC), Jaccard similarity index (JSI) For classification and disease detection performance, Precision, Recall, F-measure, and Accuracy are used as evaluation indicators.

Subjective evaluation refers to visual inspection of the denoised images, while objective evaluation employs the quantitative measures above. In this case, the parameter dice similarity coefficient (DSC) is used to determine the exact amount of ratio of the available real tumor or

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

lesion and available nontumor or non-lesion pixels to the anticipated tumor or lesion and nontumor pixels and is computed using equation 27 as follows:

$$\mathrm{DSC} = \frac{(2\mathrm{TP})}{(\mathrm{FP}+2\mathrm{TP}+\mathrm{FN})} \times 100 \tag{27}$$

The Jaccard similarity index (JSI) is used to calculate the percentage of the similarity amid actual tumor or lesion pixels in the region of interest and the number of anticipated tumor pixels and is computed as per the standard equation 28 as follows.

$$\mathrm{JSI} = \frac{(\mathrm{TP})}{(\mathrm{TP}+\mathrm{FN}+\mathrm{FP})} \times 100 \tag{28}$$

For classification, Precision, Recall, F-measure, Accuracy and Error Rate are derived using true positive (TP), false positive (FP), true negative (TN), and false negative (FN) samples:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{29}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{30}$$

$$\mathrm{F-measure} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{31}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{32}$$

$$\mathrm{Error\ Rate} = 100 - \mathrm{Accuracy} \tag{33}$$

$$\mathrm{T_A} = \frac{\mathrm{T}}{\mathrm{N}} \tag{34}$$

where T denotes the total detection time for the verification set, and N represents the total number of samples within the verification set. In addition, the proposed model achieves the lowest average diagnostic time, denoted as $(T_A(\mathrm{ms}))$.

3.9. COMPARISON WITH TRADITIONAL SEGMENTATION METHODS

Table 5 presents the comparison between traditional segmentation techniques and the proposed model, evaluated across six classes using both quantitative and qualitative analyses. The quantitative evaluation employs key metrics such as Dice Similarity Coefficient (DSC) and Jaccard Similarity Index (JSI) to numerically assess the quality of the segmented images. Four segmentation methods—MTS [23], K-means [29], MBIRCH [30], and WHOAMTS—were compared. In addition, a qualitative evaluation is performed by visually analyzing the segmented outputs, providing an intuitive understanding of their perceptual quality and segmentation accuracy

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

TABLE 5. QUALITY EVALUATION METRICS VS. IMAGE SEGMENTATION METHODS

Methods/ Images	**Bacterial spot**		**Early blight**		**Leaf Mold**	
	DSC	**JSI**	**DSC**	**JSI**	**DSC**	**JSI**
MTS	0.6802	0.7441	0.6722	0.7022	0.7062	0.8222
K means	0.7207	0.8012	0.7025	0.7844	0.7667	0.8444
MBIRCH	0.8554	0.9245	0.7845	0.8645	0.8814	0.8945
WHOAMTS	0.9072	0.9567	0.8902	0.9222	0.9232	0.9422
Methods/ Images	**Septoria leaf spot**		**Leaf Curl Virus**		**Healthy**	
	DSC	**JSI**	**DSC**	**JSI**	**DSC**	**JSI**
MTS	0.6932	0.6962	0.7262	0.8422	0.7072	0.8422
K means	0.7245	0.7567	0.7967	0.8644	0.7477	0.8744
MBIRCH	0.8255	0.8714	0.9024	0.9145	0.8764	0.9145
WHOAMTS	0.9112	0.9132	0.9432	0.9622	0.9322	0.9622

Received: August 08 2025

International Journal of Applied Mathematics
Volume 38 No. 12s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

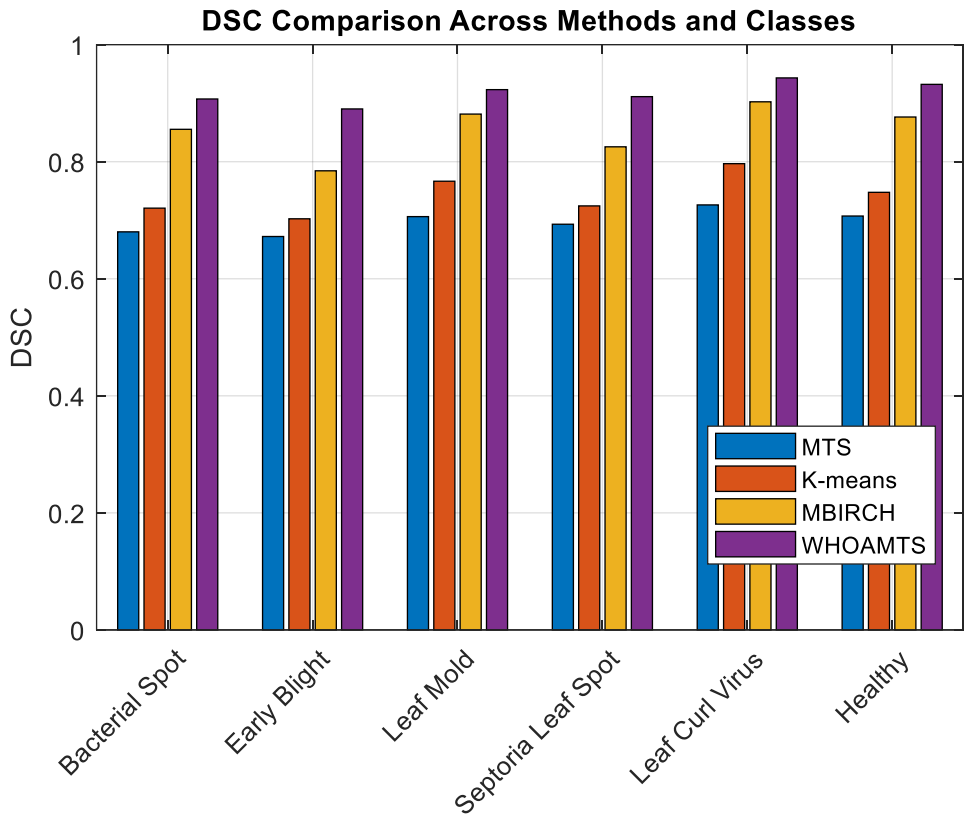


FIGURE 6. DSC COMPARISON VS. SEGMENTATION METHODS

Figure 6 shows the Dice Similarity Coefficient (DSC) values for four segmentation methods—MTS, K-means, MBIRCH, and WHOAMTS. For Bacterial Spot, the DSC values are 0.6802, 0.7207, 0.8554, and 0.9072, respectively; for Early Blight, they are 0.6722, 0.7025, 0.7845, and 0.8902; for Leaf Mold, 0.7062, 0.7667, 0.8814, and 0.9232; for Septoria Leaf Spot, 0.6932, 0.7245, 0.8255, and 0.9112; for Leaf Curl Virus, 0.7262, 0.7967, 0.9024, and 0.9432; and for Healthy leaves, 0.7072, 0.7477, 0.8764, and 0.9322. It is evident that WHOAMTS consistently achieves the highest DSC values across all classes, indicating superior segmentation accuracy, while MBIRCH shows moderate performance and MTS and K-means have comparatively lower DSC, especially for Early Blight and Septoria Leaf Spot.

. Figure 7 illustrates the Jaccard Similarity Index (JSI) values for the same classes and methods. The JSI values for Bacterials Spot are 0.7441, 0.8012, 0.9245, and 0.9567; for Early Blight, 0.7022, 0.7844, 0.8645, and 0.9222; for Leaf Mold, 0.8222, 0.8444, 0.8945, and 0.9422; for Septoria Leaf Spot, 0.6962, 0.7567, 0.8714, and 0.9132; for Leaf Curl Virus, 0.8422, 0.8644, 0.9145, and 0.9622; and for Healthy, 0.8422, 0.8744, 0.9145, and 0.9622. Similar to DSC, WHOAMTS outperforms all other methods in JSI across all classes, demonstrating higher overlap with the ground truth segmentation. MBIRCH consistently performs better than K-means and MTS, which generally show lower segmentation performance. Overall, these results confirm the robustness and accuracy of WHOAMTS in segmenting both diseased and healthy tomato leaf regions

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

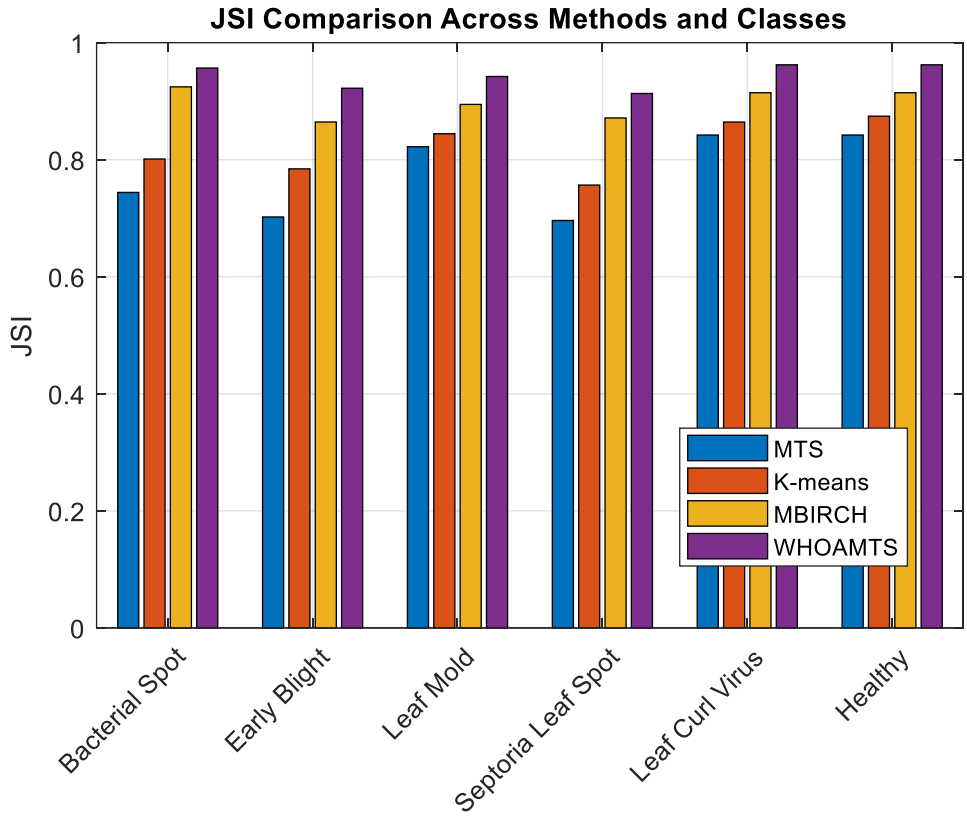


FIGURE 7. JSI COMPARISON VS. SEGMENTATION METHODS

TABLE 6. EVALUATION RESULTS VS. LEAF DISEASE DETECTION METHODS

Methods	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)	Error Rate (%)	Diagnostic time (ms)
B-ARNet	85.89	84.09	84.98	88.14	11.86	40.62
DIMPCNET	88.08	86.65	87.36	90.12	9.88	36.49
M−AORANet	89.57	88.63	89.09	91.63	8.37	33.75
MSDAIV3	91.76	91.19	91.47	93.52	6.48	30.47
MAMViT	95.53	95.50	95.51	95.00	5.00	27.71

Figure 8 depicts the comparative analysis of B-ARNet, DIMPCNET, M−AORANet, MSDAIV3, and the proposed MAMViT model in terms of Precision and Recall metrics. The Precision values obtained by B-ARNet, DIMPCNET, M−AORANet, MSDAIV3, and MAMViT are 85.89%, 88.08%, 89.57%, 91.76%, and 95.53%, respectively. Similarly, the Recall values achieved by these models are 84.09%, 86.65%, 88.63%, 91.19%, and 95.50%,

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

respectively. When compared with the proposed MAMViT framework, B-ARNet, DIMPCNET, and M−AORANet exhibit 9.64%, 7.45%, and 5.96% lower Precision values, respectively, whereas MSDAIV3 shows a relatively smaller difference of 3.77%. In terms of Recall, B-ARNet, DIMPCNET, and M−AORANet demonstrate 11.41%, 8.85%, and 6.87% lower performance compared to MAMViT, while MSDAIV3 records a marginal reduction of 4.31%. These results clearly indicate that the proposed MAMViT model consistently achieves superior Precision and Recall performance, highlighting its enhanced capability to accurately identify disease-affected regions while minimizing false positive and false negative predictions.

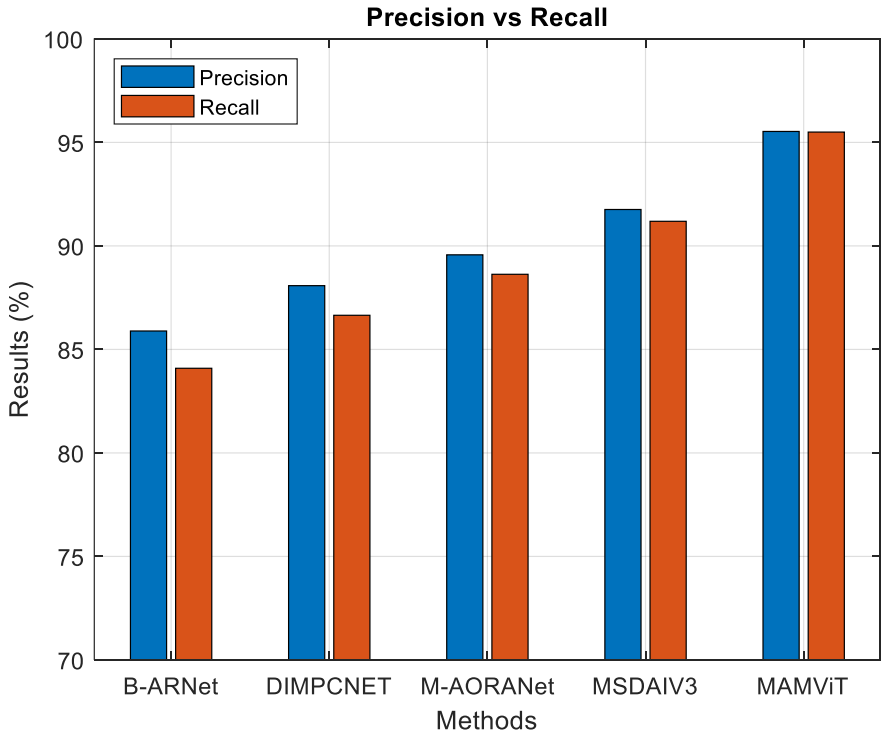


FIGURE 8. PRECISION AND RECALL COMPARISON VS. DETECTION METHODS

Received: August 08 2025

International Journal of Applied Mathematics

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

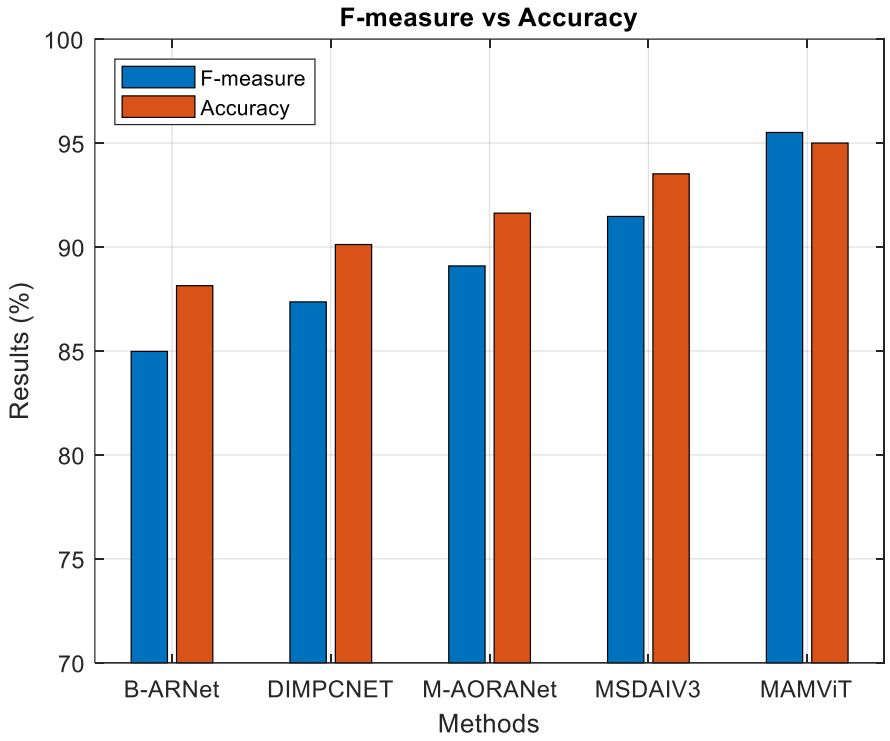


FIGURE 9. F-MEASURE AND ACCURACY COMPARISON VS. DETECTION METHODS

Figure 9 illustrates the comparative performance of B-ARNet, DIMPCNET, M−AORANet, MSDAIV3, and the proposed MAMViT framework based on F-measure and Accuracy metrics. The F-measure values achieved by B-ARNet, DIMPCNET, M−AORANet, MSDAIV3, and MAMViT are 84.98%, 87.36%, 89.09%, 91.47%, and 95.51%, respectively. Correspondingly, the Accuracy values obtained are 88.14%, 90.12%, 91.63%, 93.52%, and 95.00%, respectively. Compared with the proposed MAMViT model, B-ARNet, DIMPCNET, and M−AORANet achieve 10.53%, 8.15%, and 6.42% lower F-measure values, respectively, while MSDAIV3 exhibits a comparatively smaller difference of 4.04%. In terms of Accuracy, B-ARNet, DIMPCNET, and M−AORANet show 6.86%, 4.88%, and 3.37% lower results than MAMViT, respectively, whereas MSDAIV3 records a marginal reduction of 1.48%. These comparative results confirm that the proposed MAMViT framework significantly outperforms existing methods in both segmentation effectiveness and disease classification accuracy. This improvement can be attributed to the integration of advanced denoising strategies, optimized segmentation, and attention-driven multi-scale and multi-axis feature learning mechanisms.

4. CONCLUSION

The proposed LCGAN–WHOAMTS–Inception-V3–MAMViT framework effectively enhances tomato leaf disease detection through a unified deep learning and optimization approach.

The LCGAN module improves image quality by eliminating noise and illumination distortions. WHOAMTS achieves precise disease region segmentation using adaptive weight and bio-

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

inspired optimization. Inception-V3 extracts rich multi-scale features essential for robust classification. The MAMViT model integrates multi-scale and multi-axis attention for accurate disease identification.

Evaluation using DSC,JSI, precision, recall, F-measure, and accuracy demonstrates superior performance. The system shows strong generalization under real-world agricultural imaging conditions.

This hybrid model ensures reliable, automated disease monitoring for precision farming. Future work will extend the approach to multi-crop disease detection and real-time field deployment.

REFERENCES

1. Dutta, K., Talukdar, D. and Bora, S.S., 2022. Segmentation of unhealthy leaves in cruciferous crops for early disease detection using vegetative indices and Otsu thresholding of aerial images. *Measurement*, *189*, p.110478.
2. Mattos, A.D.P., Tolentino Júnior, J.B. and Itako, A.T., 2020. Determination of the severity of Septoria leaf spot in tomato by using digital images. *Australasian Plant Pathology*, *49*(4), pp.329-356.
3. 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.
4. Ngugi, L.C., Abdelwahab, M. and Abo-Zahhad, M., 2023. A new approach to learning and recognizing leaf diseases from individual lesions using convolutional neural networks. *Information Processing in Agriculture*, *10*(1), pp.11-27.
5. Sibiya, M. and Sumbwanyambe, M., 2019. An algorithm for severity estimation of plant leaf diseases by the use of colour threshold image segmentation and fuzzy logic inference: a proposed algorithm to update a "leaf doctor" application. *AgriEngineering*, *1*(2), p.15.
6. Trivedi, N.K., Gautam, V., Anand, A., Aljahdali, H.M., Villar, S.G., Anand, D., Goyal, N. and Kadry, S., 2021. Early detection and classification of tomato leaf disease using high-performance deep neural network. *Sensors*, *21*(23), p.7987.
7. Billah, M.M., Sultana, A., Sad Aftab, R., Ahmed, M.M. and Shorif Uddin, M., 2024. Leaf disease detection using convolutional neural networks: a proposed model using tomato plant leaves. *Neural Computing and Applications*, *36*(32), pp.20043-20053.
8. Tian, X., Meng, X., Wu, Q., Chen, Y. and Pan, J., 2022. Identification of tomato leaf diseases based on a deep neuro-fuzzy network. *Journal of The Institution of Engineers (India): Series A*, *103*(2), pp.695-706.:
9. Wu, J., Wen, C., Chen, H., Ma, Z., Zhang, T., Su, H. and Yang, C., 2022. DS-DETR: A model for tomato leaf disease segmentation and damage evaluation. *Agronomy*, *12*(9), p.2023.
10. Karthik, R., Hariharan, M. and Menaka, R., 2020. Attention embedded residual CNN for disease detection in tomato leaves. *Applied Soft Computing*, *86*, p.105933.

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

11. Ahmad, I., Hamid, M., Yousaf, S., Shah, S.T. and Ahmad, M.O., 2020. Optimizing pretrained convolutional neural networks for tomato leaf disease detection. *Complexity*, *2020*(1), p.8812019.
12. Ni, J., Zhou, Z., Zhao, Y., Han, Z. and Zhao, L., 2023. Tomato leaf disease recognition based on improved convolutional neural network with attention mechanism. *Plant Pathology*, *72*(7), pp.1335-1344.
13. Qi, J., Liu, X., Liu, K., Xu, F., Guo, H., Tian, X., Li, M., Bao, Z. and Li, Y., 2022. An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease. *Computers and electronics in agriculture*, *194*, p.106780.
14. Wang, Y., Zhang, P. and Tian, S., 2024. Tomato leaf disease detection based on attention mechanism and multi-scale feature fusion. *Frontiers in Plant Science*, *15*, p.1382802.
15. Sunil, C.K. and Jaidhar, C.D., 2023. Tomato plant disease classification using multilevel feature fusion with adaptive channel spatial and pixel attention mechanism. *Expert Systems with Applications*, *228*, p.120381.
16. Chowdhury, M.E., Rahman, T., Khandakar, A., Ayari, M.A., Khan, A.U., Khan, M.S., Al-Emadi, N., Reaz, M.B.I., Islam, M.T. and Ali, S.H.M., 2021. Automatic and reliable leaf disease detection using deep learning techniques. *AgriEngineering*, *3*(2), pp.294-312.
17. Zhao, S., Peng, Y., Liu, J. and Wu, S., 2021. Tomato leaf disease diagnosis based on improved convolution neural network by attention module. *Agriculture*, *11*(7), p.651.
18. Zhou, H., Fang, Z., Wang, Y. and Tong, M., 2023. Image Generation of Tomato Leaf Disease Identification Based on Small-ACGAN. *Computers, Materials & Continua*, *76*(1).
19. Deng, H., Luo, D., Chang, Z., Li, H. and Yang, X., 2021. Rahc_gan: A data augmentation method for tomato leaf disease recognition. *Symmetry*, *13*(9), p.1597.
20. Wang, S.H., Wu, X., Zhang, Y.D., Tang, C. and Zhang, X., 2020. Diagnosis of COVID-19 by wavelet Renyi entropy and three-segment biogeography-based optimization. *International Journal of Computational Intelligence Systems*, *13*(1), pp.1332-1344.
21. Rashmi, P. and Gomathi, R., 2024. Optimized deep learning frameworks for the medical image transmission in IoMT environment. *J. Smart Internet Things*, *2024*(2), pp.148-165.
22. Amiri, M.H., Mehrabi Hashjin, N., Montazeri, M., Mirjalili, S. and Khodadadi, N., 2024. Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm. *Scientific Reports*, *14*(1), p.5032.
23. Ramesh Babu, P., Srikrishna, A. and Gera, V.R., 2024. Diagnosis of tomato leaf disease using OTSU multi-threshold image segmentation-based chimp optimization algorithm and LeNet-5 classifier. *Journal of Plant Diseases and Protection*, *131*(6), pp.2221-2236.

Received: August 08 2025

Volume 38 No. 12s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

24. Thomkaew, J. and Intakosum, S., 2022. Improvement classification approach in tomato leaf disease using modified visual geometry group (vgg)-inceptionv3. *International Journal of Advanced Computer Science and Applications*, *13*(12).
25. Fan, H., Xiong, B., Mangalam, K., Li, Y., Yan, Z., Malik, J. and Feichtenhofer, C., 2021. Multiscale vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6824-6835).
26. Huang, H., Chen, Z., Zou, Y., Lu, M., Chen, C., Song, Y., Zhang, H. and Yan, F., 2024. Channel prior convolutional attention for medical image segmentation. *Computers in Biology and Medicine*, *178*, p.108784.
27. Li, Y., Wu, C.Y., Fan, H., Mangalam, K., Xiong, B., Malik, J. and Feichtenhofer, C., 2022. Mvitv2: Improved multiscale vision transformers for classification and detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4804-4814).
28. Li, X., Yu, J., Jiang, S., Lu, H. and Li, Z., 2023. Msvit: training multiscale vision transformers for image retrieval. *IEEE Transactions on Multimedia*, *26*, pp.2809-2823.
29. Kumari, C.U., Prasad, S.J. and Mounika, G., 2019, March. Leaf disease detection: feature extraction with K-means clustering and classification with ANN. In *2019 3rd international conference on computing methodologies and communication (ICCMC)* (pp. 1095-1098). IEEE.
30. Jayanthi, V. and Kanchana, M., 2024, March. Modified BIRCH and Hybrid Deep Framework for Tomato Leaf Disease Segmentation and Classification. In *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 149-160). IEEE.

Received: August 08 2025