RESEARCH PAPER

Tomato Leaf Disease Detection Using IGWO Feature Selection and AMO-Optimized Multi-Scale Attention Multi-Axis Vision Transformer

N.S. Tamil Ilakkiya[1], Dr. P.M. Gomathi[2]
[1]Ph.D Research Scholar, Department of Computer Science, PKR Arts College for Women, Gobi, Bharathiyar University, Coimbatore, Tamil Nadu, India
[2]Associate Professor, Dean and Supervisor, Department of Computer Science, PKR Arts College for Women, Gobi, Bharathiyar University, Coimbatore, Tamil Nadu, India

ABSTRACT: The integration of artificial intelligence with agricultural science has significantly advanced precision farming by enabling automated monitoring and diagnosis of crop health. In this study, a hybrid deep learning and statistical-based framework is proposed for efficient tomato leaf disease detection and classification. The system begins by pre-processing the input leaf images using a Lagrange Conditional Generative Adversarial Network (LCGAN), which enhances image contrast and reduces background noise for improved segmentation accuracy. The Weight Hippopotamus Optimization Multi-Threshold Segmentation (WHOAMTS) technique is then employed to effectively separate diseased areas from healthy regions. Subsequently, a hybrid Statistical Feature–Based Inception-V3 model is utilized for feature extraction, combining deep convolutional representations with handcrafted statistical features, including GLCM-based texture descriptors, colour moments, and geometric shape parameters. This integration ensures comprehensive capture of both high-level semantic information and fine-grained spatial details. The extracted hybrid features are further refined using the Improved Grey Wolf Optimization (IGWO) algorithm to select the most discriminative feature subset, thereby enhancing classification performance while reducing redundancy. Finally, the optimized features are classified using an Optimized Multi-Scale Attention Multi-Axis Vision Transformer (OMAMViT), where the transformer's attention weights and hyperparameters are fine-tuned using the Animal Migration Optimization (AMO) algorithm. The proposed hybrid approach leverages statistical–deep feature fusion, intelligent feature selection, and metaheuristic transformer optimization to achieve superior accuracy, faster convergence, and robust generalization across multiple tomato leaf disease categories under diverse imaging conditions.

Keywords: Tomato leaf disease detection, Lagrange Conditional GAN (LCGAN), Weight Hippopotamus Optimization Multi-Threshold Segmentation (WHOAMTS), Improved Grey Wolf Optimization (IGWO), Optimized Multi-Scale Attention Multi-Axis Vision Transformer (OMAMViT), Animal Migration Optimization (AMO).

How to cite this article: Tamil Ilakkiya NS, Gomathi PM. Tomato Leaf Disease Detection Using IGWO Feature Selection and AMO-Optimized Multi-Scale Attention Multi-Axis Vision Transformer. Int J Drug Deliv Technol. 2026;16(13s): 950-974. DOI: 10.25258/ijddt.16.13s.108

1. INTRODUCTION

Tomato cultivation holds significant importance in global agriculture, providing essential nutrients and supporting major economic activities across several countries. However, tomato plants are highly vulnerable to a broad spectrum of diseases such as early blight, Septoria leaf spot, mosaic virus, and bacterial speck, which substantially diminish yield quality and productivity[1]. Traditional disease diagnosis largely depends on expert visual inspection, which is subject to human bias and requires considerable time and experience, especially when symptoms are subtle or overlapping. In large farmlands, manual monitoring becomes practically infeasible, making automated, intelligent disease detection an essential requirement for modern precision agriculture. With advancements in computer vision and deep learning, image-based plant disease diagnosis has evolved into a promising alternative that offers scalability, consistency, and real-time prediction capabilities [2].

Although deep learning models, particularly convolutional neural networks (CNNs), have shown significant potential for plant disease identification, several challenges limit their real-world adaptability. Variations in image quality, inconsistent illumination, shadows, background interference, and occlusion often degrade the segmentation and classification accuracy of standard CNN-based approaches. Moreover, the morphological similarity between healthy and diseased regions, especially in early infection stages, requires models capable of capturing both high-level semantic features and fine-grained textural patterns [3,4]. Without robust pre-processing, segmentation, and feature refinement steps, even advanced deep learning architectures struggle to deliver consistent performance across diverse imaging conditions, which limits their deployment in uncontrolled field environments.

In our earlier research, a segmentation- and deep-learning-based tomato leaf disease detection system was developed. However, a careful evaluation revealed several limitations that motivated the development of a more robust framework [5]. The segmentation module lacked resilience to background noise and lighting variations, causing partial extraction of disease spots and affecting subsequent feature quality. The feature extraction component relied heavily on deep convolutional descriptors [6], which, although effective, failed to fully capture essential colour, texture, and geometric variations necessary for distinguishing closely related disease types. Additionally, the classification accuracy suffered due to the presence of redundant features and the absence of a powerful attention-based mechanism for adaptive feature weighting. These limitations highlighted the need for an improved model integrating advanced pre-processing, hybrid feature engineering, and optimization-driven classification [7,8].

To address these shortcomings, the present study proposes a comprehensive and hybrid framework that leverages both deep learning [9,10] and statistical feature analysis to produce a more discriminative and noise-resilient representation of tomato leaf diseases. The pre-processing stage employs a Lagrange Conditional Generative Adversarial Network (LCGAN), which enhances image sharpness and contrast while suppressing background elements that commonly interfere with leaf segmentation. This significantly improves the clarity of disease patterns before segmentation. Following this, the Weight Hippopotamus Optimization Multi-Threshold Segmentation (WHOAMTS) method is introduced to identify and isolate diseased regions more accurately. The combination of LCGAN and WHOAMTS ensures consistent segmentation performance, even in complex imaging scenarios involving shadows, poor lighting, or textured backgrounds.

Beyond segmentation, the proposed hybrid Statistical Feature–Based Inception-V3 module extracts a rich set of complementary features by merging deep semantic embeddings with handcrafted statistical descriptors. The handcrafted features include GLCM texture parameters, colour moments, and geometric attributes that capture subtle disease-induced variations often overlooked by deep networks alone. This multi-level feature design improves the system’s ability to differentiate between disease categories with highly similar visual characteristics. However, hybrid features introduce redundancy and high dimensionality, which can negatively impact computation time and classifier performance. Hence, the Improved Grey Wolf Optimization (IGWO) algorithm is employed to intelligently select the most relevant feature subset while reducing noise and minimizing computational overhead. Finally, to further enhance classification performance, an Optimized Multi-Scale Attention Multi-Axis Vision Transformer (OMAMViT) is implemented, where transformer hyperparameters and attention weights are fine-tuned using the Animal Migration Optimization (AMO) algorithm. This strategic optimization empowers the transformer to focus on the most discriminative regions in the input data, enabling superior learning of disease-specific patterns across scales and orientations. The synergistic integration of pre-processing, segmentation, hybrid feature extraction, feature selection, and transformer-based classification results in a highly efficient, accurate, and generalizable system for tomato leaf disease detection. The proposed model demonstrates strong resilience in diverse environmental conditions and holds substantial potential for real-time agricultural disease monitoring applications.

2. LITERATURE REVIEW

Cheemaladinne and Reddy [2024] [11] presented an innovative VARMAx-CNN-GAN (Vector Autoregressive Moving Average with Exogenous Inputs–Convolutional Neural Network–Generative Adversarial Network) integration framework for tomato leaf disease detection. In this model, the CNN extracts features and performs early disease classification, while the GAN generates synthetic tomato leaf images to enlarge the dataset and improve generalization. The method was tested on a custom tomato leaf dataset containing real and GAN-generated images. Performance was evaluated using accuracy, precision, recall, and AUC (Area Under the Curve), demonstrating superior classification efficiency and faster disease identification compared to traditional approaches.

Khan et al. [2024] [12] proposed a robust methodology combining GLCM (Gray Level Co-occurrence Matrix) and SIFT (Scale-Invariant Feature Transform) feature extraction with an SVM (Support Vector Machine) classifier. The system was trained using a dataset of 2700 tomato leaf images comprising nine disease categories. Evaluation based on accuracy, precision, recall, and comparative analysis with other state-of-the-art methods confirmed the effectiveness and reliability of this approach for multiclass tomato disease detection.

Paulchamy et al. [2025] [13] developed an improved system for tomato crop disease classification, where preprocessing was carried out using a median filter and image enhancement was based on contour detection. Morphological opening and closing operations supported noise removal, followed by FCM (Fuzzy C-Means) clustering for segmentation. GLCM texture features were extracted and classified using an ANN (Artificial Neural Network). Experiments on publicly available and field-collected tomato datasets showed improved accuracy and precision.

Luna-Benoso et al. [2021] [14] introduced a three-stage tomato disease detection framework consisting of leaf segmentation based on pixel color values, feature extraction using color moments and GLCM, and classification using SVM, K-NN (K-Nearest Neighbors), and MLP (Multilayer Perceptron). A consensus decision rule was employed, selecting the decision shared by at least two classifiers. Evaluation using accuracy and cross-validation demonstrated competitive performance.

Negi et al. [2024] [15] implemented MGWO (Modified Grey Wolf Optimization) for feature selection and combined it with standard classifiers for tomato disease classification. Using a tomato plant dataset, MGWO was compared against GA (Genetic Algorithm) and PSO (Particle Swarm Optimization) and outperformed both in terms of classification accuracy and optimization efficiency.

Gülcü [2022] [16] proposed the IAMO-MLP (Improved Animal Migration Optimization–Multilayer Perceptron) hybrid algorithm to enhance neural network training. The algorithm was evaluated using benchmark datasets including XOR, Balloon, Iris, Breast Cancer, and Heart, along with a real-world dataset. Performance was measured using MSE (Mean Squared Error), classification accuracy, and the Friedman statistical test, proving improved convergence and robustness.

Tu et al. [2022] [17] introduced MaxViT (Multi-Axis Vision Transformer), a hierarchical vision backbone that integrates attention mechanisms with convolutional operations. The architecture was evaluated on large-scale datasets such as ImageNet-1K and ImageNet-21K, achieving state-of-the-art Top-1 accuracy. MaxViT also delivered strong performance on downstream tasks including object detection (e.g., the COCO dataset) and demonstrated high-quality generative modeling capabilities.

3. PROPOSED METHODOLOGY

The proposed methodology introduces an advanced hybrid framework integrating deep learning, statistical feature engineering, and intelligent optimization for tomato leaf disease detection. Initially, a Lagrange Conditional GAN (LCGAN) enhances image quality by improving contrast and reducing background noise. The refined images are segmented using the Weight Hippopotamus Optimization Multi-Threshold Segmentation (WHOAMTS) approach to accurately isolate diseased regions. A hybrid Statistical Feature–Based Inception-V3 model then extracts deep semantic and handcrafted statistical descriptors. Redundant features are eliminated through Improved Grey Wolf Optimization (IGWO), ensuring a compact and discriminative feature subset. Finally, disease classification is performed using an AMO-optimized Multi-Scale Attention Multi-Axis Vision Transformer (OMAMViT), enabling precise and robust identification across multiple disease categories. Overall process of proposed system is illustrated in figure 1.

3.1 DATASET COLLECTION

In this study, the tomato leaf dataset was obtained from real-time sources and publicly accessible repositories to support image processing and computer vision tasks. The original dataset consists of nine categories representing various healthy and diseased tomato leaf conditions. 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. All images are provided in the RGB colour format with an initial resolution of 256×256 pixels, which was resized to 224×224 pixels to maintain uniformity during model training. Background regions were removed prior to processing to enhance the visibility of disease symptoms and improve feature extraction accuracy. A total of 10,239 tomato leaf images were utilized for experimentation, 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. The dataset includes variations in disease severity, leaf texture, and colour distribution, ensuring that the model learns discriminative features effectively. Its comprehensive coverage across multiple disease categories makes it well-suited for evaluating the performance and generalization capability of the proposed detection framework.

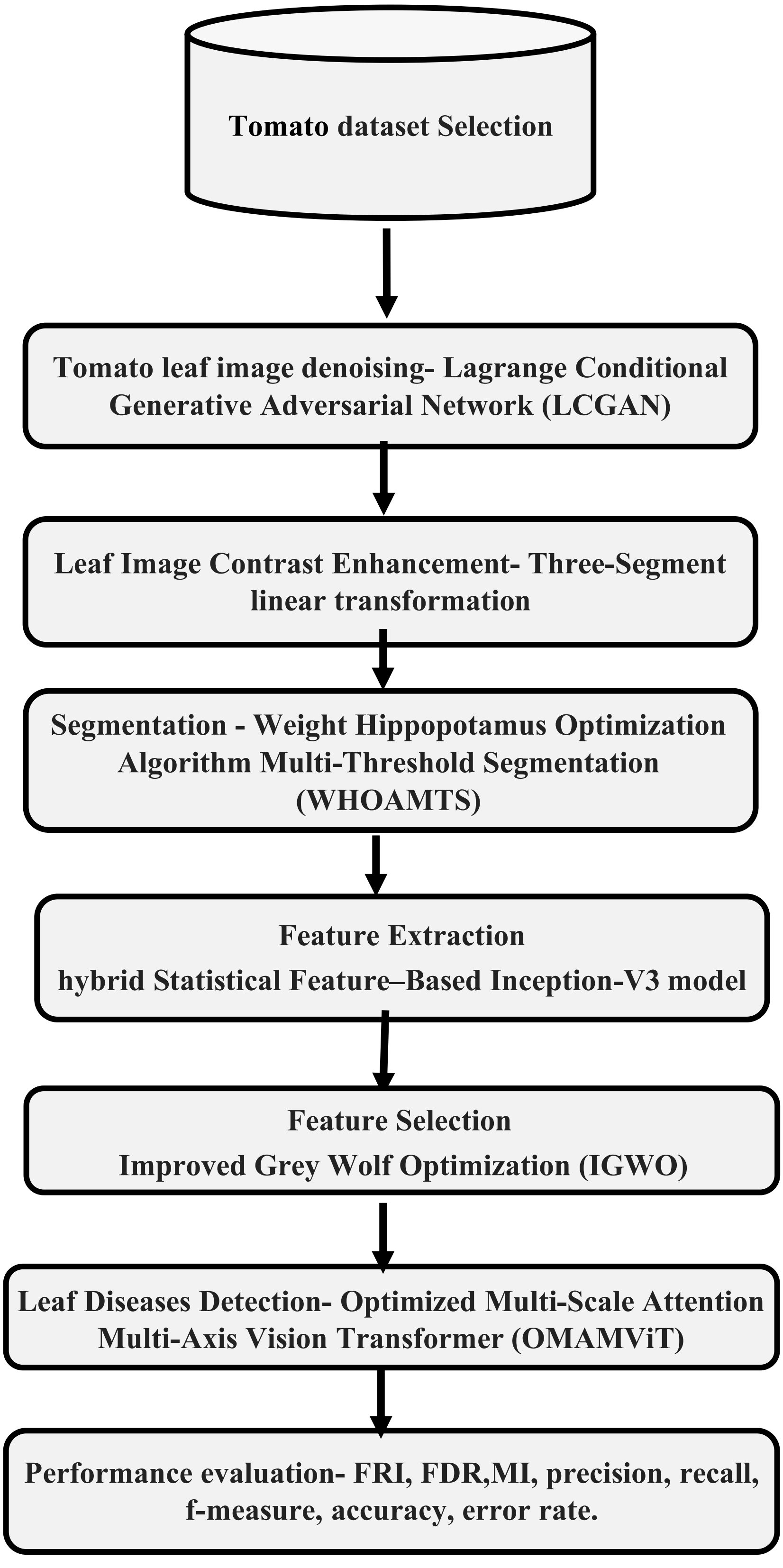


FIGURE 1. PROPOSED TOMATO LEAF DISEASES DETECTION MODEL

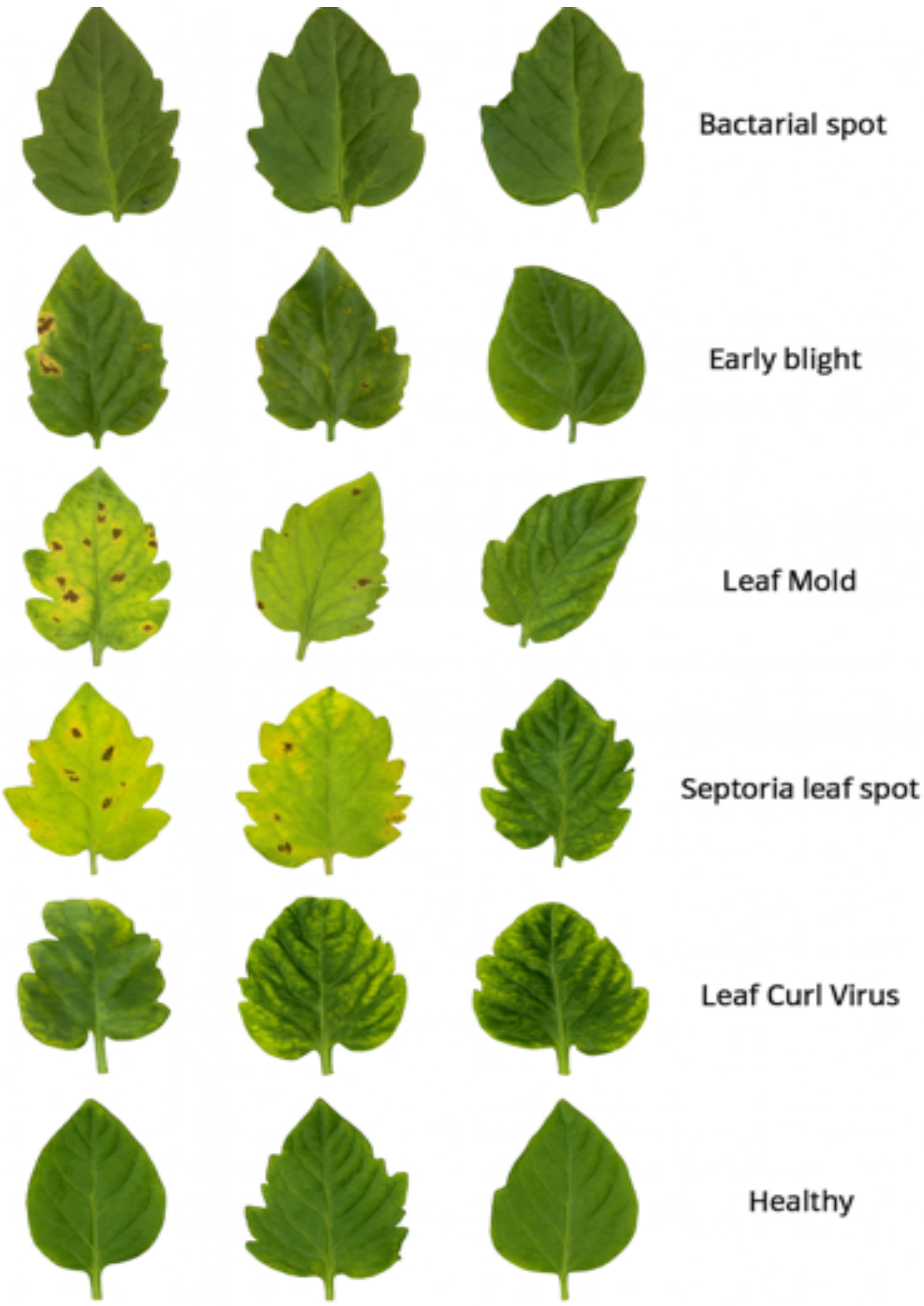


FIGURE 2. SAMPLE OF TOMATO LEAVES FROM THE REAL TIME DATASET

As detailed in Table 1, the tomato leaf dataset was systematically divided into three subsets to support model development: 70% for training, 20% for testing, and 10% for validation. All leaf images were collected under natural field conditions using a digital camera to capture real-world variations. Prior to further processing, an image enhancement technique was applied to eliminate noise, improve contrast, and refine the visual quality of the leaf samples, ensuring better feature extraction and reliable classification. This structured dataset preparation helps maintain consistency across the learning phases and contributes to achieving robust and accurate disease prediction performance.

TABLE 1. DISTRIBUTION OF TOMATO LEAF DISEASE IMAGES ACROSS DATA SPLITS

Class Category	Total Samples	Training Set	Testing Set	Validation Set
Bacterial Spot	1,702	1,191	340	171
Early Blight	800	560	160	80
Leaf Mold	762	535	152	75
Septoria Leaf Spot	1,417	992	283	142
Leaf Curl Virus	4,286	3,000	857	429
Healthy Leaves	1,272	891	254	127

3.2. LAGRANGE CONDITIONAL GAN–BASED IMAGE DENOISING

The Lagrange Conditional Generative Adversarial Network (LCGAN) is employed in this work to remove noise from tomato leaf images while preserving subtle structural information essential for accurate disease recognition. This model incorporates a Lagrange-based constraint that enforces strong consistency between the denoised image and its corresponding ground truth, while the adversarial learning mechanism enhances the perceptual realism of the output. The LCGAN framework consists of a densely connected generator, a multi-scale discriminator, and a refined perceptual loss unit [18]. The generator is designed as a symmetric deep CNN with dense blocks, transition layers, and skip connections that promote effective feature reuse and maintain important spatial characteristics during reconstruction. In parallel, the discriminator operates across multiple scales to assess both local textures and global patterns, enabling it to reliably distinguish real clean images from generated ones. To guide the model toward producing visually coherent results, a refined loss function integrates pixel-level Euclidean loss, perceptual similarity loss, and adversarial loss, each weighted using Lagrange multipliers to achieve an optimal balance. This combined strategy ensures that the denoised images are smooth, structurally accurate, and visually enhanced. After the LCGAN stage, a final contrast enhancement step is applied to further improve clarity before segmentation and subsequent analysis.

3.3 THREE-SEGMENT LINEAR TRANSFORMATION BASED CONTRAST ENHANCEMENT

Three-segment linear transformation is employed to enhance the contrast of an image effectively. In this approach, the input denoised image is adjusted in three distinct intensity ranges, each with a different slope to control the degree of contrast enhancement. The first segment targets the lower intensity values, applying a gentle amplification, while the middle segment adjusts the mid-range intensities to further boost the image details. The final segment focuses on higher intensity values, ensuring the bright regions are adequately enhanced without saturation. The points where these slope changes occur are carefully selected so that the overall contrast of the image is improved, resulting in a visually sharper and more detailed output.

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

The Weight Hippopotamus Optimization Algorithm Multi-Threshold Segmentation (WHOAMTS) is a bio-inspired optimization method designed to achieve precise and efficient segmentation of tomato leaf disease images . The algorithm models the natural behaviors of hippopotamuses—such as exploration, defensive movement, and escape—allowing it to maintain a balanced trade-off between global search and localized refinement [19]. The inclusion of an Adaptive Weight Mechanism (AWM) further strengthens this balance by dynamically regulating the intensity of exploration and exploitation during the search process. This adaptive behavior improves convergence stability and segmentation accuracy, particularly in images where disease patterns exhibit significant variations in color distribution and texture complexity.

Initialization and Exploration Phase

The optimization process begins by initializing a population of hippopotamuses, each representing a candidate solution composed of multiple threshold values. These thresholds are randomly selected within the permissible intensity range of the image to ensure adequate diversity at the start of the search:

$$x_{ij} = lb_j + r \times \left(ub_j - lb_j\right) \quad (1)$$

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 1). In the tomato leaf disease dataset, these bounds correspond to pixel intensity limits within RGB or grayscale channels.During exploration, hippopotamuses move across broad regions of the search space to identify potential threshold sets. Their movement is guided partly by the dominant hippopotamus (global best solution), while randomness preserves population diversity. The adaptive exploration weight y_1 controls the magnitude of movement, enabling dynamic adjustment as iterations progress:

$$x_{ij}(t+1) = x_{ij}(t) + y_1 \times \left(D_{\text{hippo}} - I_1 \times x_{ij}(t)\right) \quad (2)$$

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 2). This phase 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) and Predator Defense Phase

The AWM plays a key role in controlling the algorithm’s search pattern. At initial iterations, a high weight value encourages broad exploration, while in later iterations, the weight gradually decreases to enhance local exploitation:

$$\text{wt} = \omega_{max;} - \frac{(\omega_{max;-}\omega_{min}) \times\ t}{t_{max}} \quad (3)$$

Here, $\omega_{max;-}\omega_{min}$ represent the maximum and minimum weight values, respectively, and t_{max} denotes the total number of iterations (Equation 3). 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.

When the population risks stagnation, a predator defense mechanism is activated. In nature, hippopotamuses respond defensively to threats; in WHOAMTS, this is modeled by relocating some individuals to unexplored zones. This diversification step generates new candidate thresholds:

$$\text{Predator}_j = lb_j + r \times \left(ub_j - lb_j\right) \quad (4)$$

This mechanism introduces fresh solutions into the search space, enabling the algorithm to avoid local optima and increasing robustness in complex segmentation tasks.

Escape Phase (Exploitation) and Fitness Evaluation

If defensive relocation does not yield improvement, the algorithm triggers an escape phase that emphasizes local optimization. Hippopotamuses move toward safer and more promising regions using locally defined bounds:

$$x_{ij}(t+1) = x_{ij}(t) + r \times \left(lb_j^{\text{local}} + s_1 \times| \left(ub_j^{\text{local}} - lb_j^{\text{local}}\right)\right) \quad (5)$$

Here, ub_j^{local} and lb_j^{local} define the local search boundaries (Equation 5) , while s_1 is the step-size control parameter. This refined search captures subtle lesion boundaries and enhances segmentation precision. Each solution (set of thresholds) is evaluated using a segmentation quality metric such as Otsu’s variance measure (figure 3). The optimization iteratively updates all hippopotamuses until the dominant threshold set is identified. These optimal thresholds are then applied to the tomato leaf image to produce multi-level segmented regions suitable for feature extraction and classification using Inception-V3 and MAMViT models.

Algorithm: WHOAMTS – Weight Hippopotamus Optimization Algorithm for Multi-Threshold Segmentation

Step 1: Begin by providing the tomato leaf image I; convert it to grayscale when required.

Step 2: Set the algorithm parameters, including the number of hippopotamuses (population size N), number of thresholds L, maximum iteration count t_max and the adaptive weight bounds w_max, w_min

Step 3: Determine the pixel intensity search limits by assigning the lower bound lb and upper bound ub based on the image's intensity range.

Step 4: Generate the initial hippopotamus population by assigning random threshold values within the defined intensity bounds.

Step 5: Compute the fitness of all candidate solutions using a suitable segmentation quality criterion, such as Otsu's between-class variance.

Step 6: For each iteration from t=1 to t_max , update the adaptive weight according to the AWM formulation.

Step 7: Identify the dominant hippopotamus, i.e., the individual with the highest fitness value.

Step 8: Execute the exploration process, allowing solutions to move toward or away from the dominant hippopotamus to survey the global search space.

Step 9: Activate the predator-defense strategy by injecting new randomly generated individuals using the predator equation to maintain solution diversity.

Step 10: Apply the escape (exploitation) mechanism, refining candidate solutions within local boundaries to enhance threshold precision.

Step 11: Re-evaluate the fitness of all updated solutions and replace the dominant hippopotamus when a superior candidate is identified.

Step 12: Continue iterating until the stopping condition is met, i.e., when t=tmax.

Step 13: Select the optimal threshold set T_opt = D_hippo, corresponding to the final dominant hippopotamus.

Step 14: Perform multi-threshold segmentation by applying T_opt to the input image to generate the segmented output regions.

Figure 3 represents the flow chart of the Weight Hippopotamus Optimization Algorithm for Multi-Threshold Segmentation algorithm.

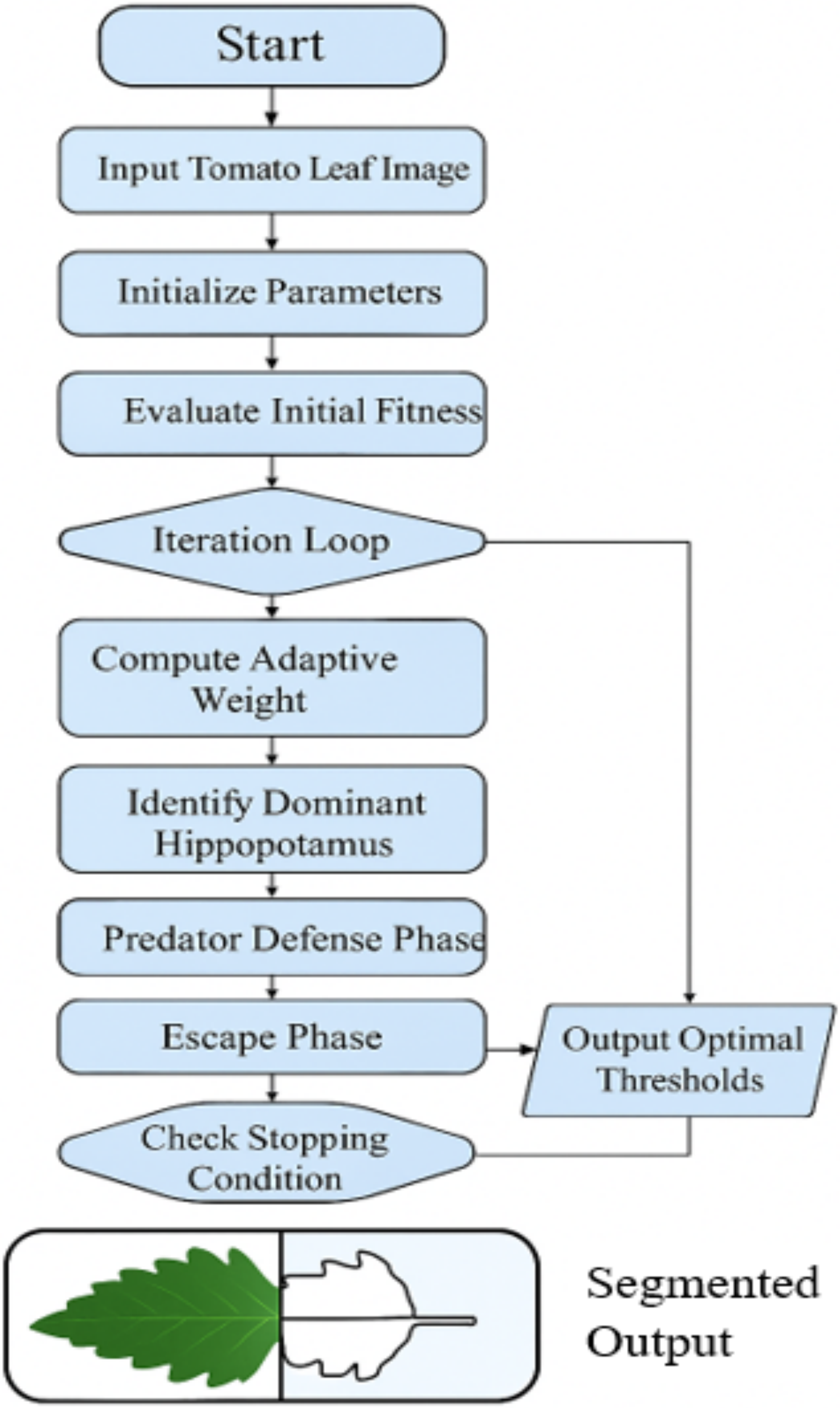


FIGURE 3. FLOWCHART OF WHOA

3.5 HYBRID STATISTICAL FEATURE–BASED INCEPTION-V3 MODEL

In this work, a hybrid approach is proposed that combines deep features extracted from the Inception-V3 network with handcrafted statistical features. Geometric, color, and texture descriptors are fused with deep semantic representations to enhance disease characterization. This hybrid feature vector serves as a robust input for the MAMViT classifier, improving the accuracy of tomato leaf disease recognition.

3.5.1 Colour Features

Initially, colour features are obtained from the enhanced and segmented tomato leaf images. Colour-based information plays a significant role in identifying infected regions, as each type of disease produces unique visual symptoms such as yellowing, browning, dark lesions, or chlorotic patches. These variations in colour provide strong discriminative cues for accurate disease recognition and classification. The enhanced tomato leaf images are used to extract colour information across all selected colour spaces. This multi–colour-space strategy ensures that maximum chromatic details related to lesion regions are captured. Each colour space is decomposed into its three respective channels (for instance, the RGB space contains red, green, and blue channels), and from each channel, four statistical measures are computed: mean, standard deviation, entropy, and skewness [20]. These statistical descriptors quantitatively represent the distribution and variation of pixel intensities within the lesion regions [21]. The fusion of all extracted features from the five colour spaces produces a robust and high-dimensional feature vector that contains richer discriminative information compared to a single colour representation.

The statistical parameters for each channel are calculated using the following equations:

- **Mean (μ):**

$$X = E(X) = \sum_{i=1}^{n} \frac{(X_i, Y_j)}{n} \tag{6}$$

- **Standard Deviation (σ):**

$$\sigma = \sqrt{E(X^2) - [E(X)]^2} \tag{7}$$

where,, $E(X^2) = \frac{1}{n}[(1)^2 + (2)^2 + \cdots (n)^2]$ (8)

$$= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} \tag{9}$$

$$E(X^2) = \frac{(n+1)(2n+1)}{6} \tag{10}$$

- **Entropy (H):**

$$Entropy = -\sum_{i,j} I(X_i, Y_j) log \frac{I(Y_j)}{I(X_i, Y_j)} \tag{11}$$

- **Skewness (S):**

$$Skewness = \frac{E(x-\bar{x})^3}{\sigma^3} \tag{12}$$

Therefore, the combined multi-colour-space features significantly improve the system's ability to distinguish and classify different tomato leaf diseases.

3.5.2 Texture Features

In this work, segmented images of tomato leaves are analyzed using Gray Level Co-occurrence Matrix (GLCM) to extract texture-based features that quantify the spatial relationship between pixels [22]. The GLCM captures the frequency of co-occurrence of pixel intensity values at a given distance and orientation in the image, providing valuable information about the texture patterns associated with disease symptoms such as spots, lesions, or discoloration [23]. For each pre-processed tomato leaf image, the GLCM is computed, and the following four statistical features are extracted:

- **Contrast:**
  Contrast measures the local variations in the gray-level co-occurrence matrix. It quantifies the intensity difference between a pixel and its neighbor over the whole image. Higher contrast values indicate significant texture variations, which are often observed in diseased regions of tomato leaves.

$$Contrast = \sum\left(|a-b|^2 \times p(a,b)\right) \tag{13}$$

- **Energy:**
  Energy, also known as Angular Second Moment (ASM), represents the uniformity of the texture. Higher energy values indicate more homogenous regions, while lower values are associated with complex textures. In tomato leaf disease detection, energy helps distinguish between healthy and diseased areas.

$$Energy = \sum p(a,b)^2 \qquad (14)$$

- **Homogeneity:**
  Homogeneity measures the closeness of the distribution of elements in the GLCM to the diagonal. Regions with less variation have higher homogeneity values. This feature is useful for identifying smooth regions in leaves, such as areas unaffected by disease.

$$Homogeneity = \frac{\sum p(a,b)^2}{1+|a-b|} \qquad (15)$$

- **Correlation:**
  Correlation evaluates the linear dependency of gray levels on neighboring pixels. It considers the mean and standard deviation of gray levels in the GLCM. Diseased regions with irregular patterns typically show distinct correlation values compared to healthy regions.

$$Correlation = \sum(a-\mu a)(b-\mu b)\frac{p(a,b)}{[\sigma a,\sigma b]} \qquad (16)$$

Here, $p(a,b)$ represents the normalized GLCM probability, a and b are pixel intensity levels, μ is the mean, and σ is the standard deviation of the gray levels.These four features—Contrast, Energy, Homogeneity, and Correlation—serve as key indicators in distinguishing healthy and diseased tomato leaves. By quantifying texture variations, they provide a robust feature set for subsequent classification using machine learning or deep learning models.

3.5.3 Geometric Features

Geometric features provide essential information about the shape, size, and spatial structure of lesions present on tomato leaves. Here eight geometric features are extracted from the segmented spot lesions to capture both global and local information, including lesion morphology, size, and position. The features include area, aspect ratio, filled area, major axis length, minor axis length, extent, perimeter, and solidity. The area represents the total number of pixels within a lesion, while the aspect ratio describes the elongation or roundness of the lesion. The filled area accounts for the lesion size after filling small holes, and the major and minor axis lengths indicate the primary and secondary orientation of the lesion, respectively. Extent measures the ratio of the lesion area to the bounding box area, providing an indication of compactness, whereas the perimeter captures the length of the lesion boundary, reflecting shape irregularities. Solidity, defined as the ratio of lesion area to its convex hull area, indicates the convexity or concavity of the lesion. These geometric features provide comprehensive information about both the global and local characteristics of the lesions. When combined with texture and color features, they significantly enhance the accuracy of tomato leaf disease classification. The final geometric feature vector has a dimension of 1×8.

3.5.4 Deep Feature Extraction Using Inception-V3

The Inception-V3 network is employed to extract deep discriminative features that characterize complex disease patterns. Each segmented leaf image is resized to 224 × 224 × 3 and input into the pretrained Inception-V3 model. This network utilizes parallel convolutional kernels of varying receptive fields (1×1, 3×3, and 5×5) to simultaneously capture fine-grained lesion textures and global structural information. The early convolutional layers extract low-level features such as edges, gradients, and color variations, while the deeper layers encode high-level semantic patterns corresponding to different disease symptoms, including Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, and Leaf Curl Virus. Inception-V3 incorporates factorized convolutions, batch normalization, and auxiliary classifiers to reduce computational complexity and mitigate overfitting. Finally, the output feature maps are extracted from the Global Average Pooling (GAP) layer, generating compact and robust feature vectors that serve as input to the subsequent classifier for multi-scale attention-based disease recognition.

3.5.5 Hybrid Feature Representation

To enhance disease classification accuracy, the deep features extracted from Inception-V3 are fused with handcrafted features, including geometric, colour, and texture descriptors. Geometric features capture the shape, size, and spatial structure of lesions, colour features encode the distribution of colour intensities, and texture features derived from the GLCM matrix provide information about the spatial relationships between pixels. By combining these feature types, the hybrid representation leverages both high-level semantic information from deep networks and low-level handcrafted descriptors, enabling more robust and discriminative disease characterization. The final hybrid feature vector integrates Inception-V3 deep features with the geometric (1×8), color (1×N), and texture (1×4) features, providing a comprehensive representation for the classifier, which performs multi-scale attention-based recognition of tomato leaf diseases with high accuracy.

3.6. FEATURE SELECTION BY IMPROVED GREY WOLF OPTIMIZATION (IGWO)

In the proposed tomato leaf disease recognition system, the Improved Grey Wolf Optimization (IGWO) algorithm is employed to select the most discriminative and relevant features from the combined feature set tuning Levy flight. The dataset initially contains a large number of extracted descriptors, including colour statistics, GLCM texture features, geometric lesion attributes, and deep features obtained from the Inception-V3 network. Since not all features contribute equally to disease classification, an effective Feature Selection (FS) mechanism is required to improve classification accuracy, reduce computational complexity, and avoid redundant information.

The Grey Wolf Optimization (GWO) algorithm simulates the leadership hierarchy and hunting behaviour of grey wolves in nature. The best solution is represented by α, followed by β and δ, while the remaining candidate solutions are denoted as ω. IGWO extends the standard GWO by incorporating Levy flight, which improves the algorithm's exploration ability and helps escape local optima during feature search [24]. In the IGWO method, the encircling and hunting behaviours of wolves are mathematically modelled to update feature selection masks. At each iteration, wolves adjust their positions in the feature space according to the estimated prey location (i.e., the best subset of features). Adaptive coefficients d and D regulate exploration and exploitation, enabling a smooth transition from global search ($|D| \geq 1$) to local refinement ($|D| < 1$). These adaptive parameters help maintain population diversity and accelerate convergence.

Hunting is supervised by α, β, and δ wolves in this system, and ω wolves are responsible for encircling the target to show enhanced resolution. The αleads the chase. The β, and δ may also occasionally engage in hunting. To put it another way, α, β, and δ compute the area around the prey's position, and other wolves occasionally report their own nearby locations.

The grey wolves end the chase by confronting the target when it begins to stop moving, as mentioned before. It lowers the value of d to accurately model reaching the prey. The following is a mathematical representation of hunting behaviour:

$$\vec{O} = \left|\vec{B}.\vec{Y}_k(m) - \vec{Y}(m)\right| \tag{17}$$

$$\vec{Y}(m+1) = \vec{Y}_k(m) + \vec{D}.\vec{O} \tag{18}$$

The location of the grey wolf (features) at iteration m is represented by Y(m), the prey's location is denoted as Y_k. Equations (19), and (20) are employed to compute the coefficient vectors $\vec{D}$ and $\vec{B}$ in this way,

$$\vec{D} = 2\vec{d}.\vec{r}_1 - d \tag{19}$$

$$\vec{B} = 2.\vec{r}_2 \tag{20}$$

Where r1, r2 are random values within [0, 1] and the coefficient vector $\vec{d}$, decreases gradually from 2 to 0 with a greater amount of iterations [25].

$$\vec{O}_\alpha = \left|\overrightarrow{C_1}.\vec{Y}_\alpha(m) - \vec{Y})\right| \tag{21}$$

$$\vec{O}_\beta = \left|\overrightarrow{C_2}.\vec{Y}_\beta(m) - \vec{Y})\right| \tag{22}$$

$$\vec{O}_\delta = \left|\overrightarrow{C_3}.\vec{Y}_\delta(m) - \vec{Y})\right| \tag{23}$$

$$\vec{Y}_1 = \vec{Y}_\alpha - \vec{D}_1.(\vec{O}_\alpha) \tag{24}$$

$$\vec{Y}_2 = \vec{Y}_\beta - \vec{D}_1.(\vec{O}_\beta) \tag{25}$$

$$\vec{Y}_3 = \vec{Y}_\delta - \vec{D}_1.(\vec{O}_\delta) \tag{26}$$

$$Y(t+1) = \frac{\vec{Y}_1+\vec{Y}_2+\vec{Y}_3}{3} \tag{27}$$

It gives information on α, β, and δ, as well as the approximate span at the present location. The final location of the wolves ω is found following the computation of distances. To get the best features, it is utilized to optimize the feature settings. Throughput is enhanced and the optimal fitness values are selected. Nevertheless, it presents computational difficulties and has problems with the best FS. The LGWO algorithm is presented in this study to enhance features and lower computing complexity in order to address the previously mentioned issue.

Levy flight is employed in this context to generate outcomes that are more effective. This method uses a more effective search according to Levy flight to avoid being caught in a local optimum while the GWO is unable to obtain the optimal results after a predetermined number of repetitions. Levy flight search improves both local and worldwide search capabilities at the same time. A set of random procedures known as Levy flights have leap sizes that correspond to the Levy probability distribution function. This is a formula of standard power-law distribution.

$$L(s) \sim |s|^{-1-\beta} \qquad (28)$$

An index is defined as $0 < \beta \leq 2$. A basic mathematical explanation of the Levy distribution can be expressed as.

$$L(s,\gamma,\mu) = \begin{cases} \sqrt{\gamma/2\pi \exp\left[-\frac{\gamma}{2(s-\mu)}\frac{1}{(s-\mu)}\right]} & if\ 0 < \mu < \propto \\ 0 & if\ s \leq 0 \end{cases} \qquad (29)$$

Here the collection of samples in this distribution is denoted by s , the scale parameter γ governs the distribution's scale, and μ stands for shift parameter or position. This component concurrently enhances local and global search performance.

In the next step, the wolves α, β, and δ, as well as ω, are marked. Involving hunting, attacking, and surrounding the target. Until the algorithm's output remains unchanged after a predetermined number of iterations and it is referred as a limiting value, this procedure is repeated. These wolves are being dispersed around the search area as a result of the Levy flight being utilized to carry out the search at this point.

$$S = \alpha \oplus levy(\beta)\ \alpha\beta\gamma\delta\delta$$

$$\vec{Y}_1 = \vec{Y}_\alpha + S, \vec{Y}_2 = \vec{Y}_\beta + S, \vec{Y}_3 = \vec{Y}_\delta + S \qquad (30)$$

$$\vec{Y}(t+1) = \frac{\vec{Y}_1 + \vec{Y}_2 + \vec{Y}_3}{3} \qquad (31)$$

This hunting and exploitation behaviour serves as the foundation for the multi-objective GWO's formulation. β is a crucial element in the Levy flight. As, β an arbitrary number among 0 and 2 is generated for every wolf in the solution. Various values of β lead to distinct outcomes. Greater jumps are caused by small values of β, whereas smaller jumps are caused by large values of β. Stated differently, greater values of β are more likely to result in jumps to uncharted territory (increased exploration) and avoid becoming trapped in local optimal situations. On the other hand, new positions close to the found solutions (greater exploitation) are triggered by small values of β. The LGWO is well-known for having excellent worldwide search capabilities.

Algorithm 2: IGWO for Optimal Tomato Leaf Feature Selection

Input: Extracted feature set (color, texture, geometric, deep features)

Output: Optimal discriminative feature subset for classifier

1. Initialize the wolf population with random feature selection vectors.
2. Compute the fitness of each feature subset (classification accuracy).
3. Identify α, β, and δ wolves (best three feature subsets).
4. **Repeat until maximum iterations:**
    - Update wolf positions using GWO encircling and hunting (Eq. 6–12).
    - Apply Levy flight to generate improved candidate solutions (Eq. 15–16).
    - Recalculate fitness for updated feature subsets.
    - Update α, β, and δ based on new fitness values.
5. Return the optimal feature subset.

IGWO efficiently reduces redundant, noisy, and irrelevant features while retaining the most informative descriptors for tomato leaf disease identification. This significantly enhances the classifier's accuracy and reduces computational time during inference.

3.7 LEAF DISAESE DETECTION BY OPTIMIZED MULTI-SCALE ATTENTION MULTI-AXIS VISION TRANSFORMER (OMAMVIT)

The overall architecture of the Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT) adopts an encoder-decoder pipeline designed to achieve effective hierarchical feature abstraction and spatial reconstruction for dense prediction tasks. The input image $X \in \mathbb{R}^{H\times W\times 3}$ is initially processed using a stem convolution to generate non-overlapping patch embeddings, producing low-level features $F_0 \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times C}$. These patch tokens are forwarded into the encoder, which consists of sequential MBConv layers and MaxViT transformer blocks. The forward pass of the network can be formally described as:

$$\hat{y} = D(\varepsilon(X)) \tag{32}$$

where $\varepsilon(\cdot)$ denotes the encoder responsible for extracting hierarchical multi-scale semantic descriptors, while $D(\cdot)$ represents the decoder used for spatial resolution recovery and final semantic output prediction [26].

Multi-Axis Attention Mechanism

The multi-axis attention module is a fundamental element of MAMViT, enabling joint modelling of both local and global contextual information. The mechanism alternates between two sequential stages: block attention, which focuses on local neighborhood feature relationships, and grid attention, which models long-range dependencies across the image. Block attention splits the feature tensor into $p \times p$ non-overlapping windows:

$$\text{Block} : (H, W, C) \rightarrow \left(\frac{HW}{p^2}, p^2, C\right) \tag{33}$$

Self-attention is computed inside each block:

$$x \leftarrow x + \text{Unblock}(\text{RelAttention}(\text{Block}(LN(x)))) \tag{34}$$

with relative position-aware attention:

$$\text{RelAttention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V \tag{35}$$

incorporate global semantics, grid attention rearranges the feature tensor into a sparse grid representation:

$$\text{Grid} : (H, W, C) \rightarrow \left(G^2, \frac{HW}{G^2}, C\right) \tag{36}$$

Attention is then computed and mapped back:

$$x \leftarrow x + \text{Ungrid}(\text{RelAttention}(\text{Grid}(LN(x)))) \tag{37}$$

followed by a feed-forward refinement:

$$\mathrm{x} \leftarrow \mathrm{x} + \mathrm{MLP}(\mathrm{LN}(\mathrm{x})) \tag{38}$$

Together, these two attention operations provide a balance between computational efficiency and global feature expressiveness.

Encoder and Feature Abstraction:

The encoder is arranged into hierarchical stages, each combining MBConv and MaxViT blocks. MBConv performs lightweight convolutional transformation and channel recalibration using squeeze-excitation (SE):

$$\mathrm{x} \leftarrow \mathrm{x} + \mathrm{Proj}(\mathrm{SE}(\mathrm{DWConv}(\mathrm{Conv}(\mathrm{Norm}(\mathrm{x}))))) \tag{39}$$

When spatial down sampling is required, stride-2 operations are introduced:

$$\mathrm{x} \leftarrow \mathrm{Proj}(\mathrm{Pool2D}(\mathrm{x})) + \mathrm{Proj}\big(\mathrm{SE}(\mathrm{DWConv}_{\downarrow}(\mathrm{Conv}(\mathrm{Norm}(\mathrm{x}))))\big) \tag{40}$$

Multi-Scale Attention Fusion and Skip Refinement:

To aggregate multi-level encoder features, a Multi-Scale Attention (MSA) [27] module is introduced:

$$\mathrm{F}_{\mathrm{MSA}} - \textstyle\sum_{\mathrm{i}=1}^{3} \alpha_{\mathrm{i}}\mathrm{A}(\mathrm{F}_{\mathrm{i}}) \tag{41}$$

where α_i are trainable softmax-normalized fusion weights, and $A(\cdot)$ corresponds to channel-spatial attention(figure 4). Channel and spatial refinement follow:

$$\begin{gathered} M_c(F) - \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))) \\ M_s(F') - \sigma\big(\text{Conv}([\text{AvgPool}(F'); \text{MaxPool}(F')])\big) \end{gathered} \tag{42}$$

leading to the refined representation:

$$F'' - M_s(F') \odot F' \tag{43}$$

Skip pathways employ Multi-Scale Channel-Spatial Attention (MCSA) to enhance semantic alignment between shallow and deep layers:

$$F_{\text{refined}} - (M_c(F_2) \odot F_0) \odot M_s(F_1) \tag{44}$$

Decoder and Output Prediction:

The decoder progressively reconstructs spatial resolution using learnable Patch Expand blocks:

$$Y_{i-1} - \text{PatchExpand}\ (Y_i) + F_{\text{refined}}^{(i)} \tag{45}$$

Finally, a softmax classifier produces dense segmentation outputs. Overall, the MAMVIT architecture unifies transformer-based long-range feature modeling with convolutional inductive priors, enabling robust multiscale representation learning [28].

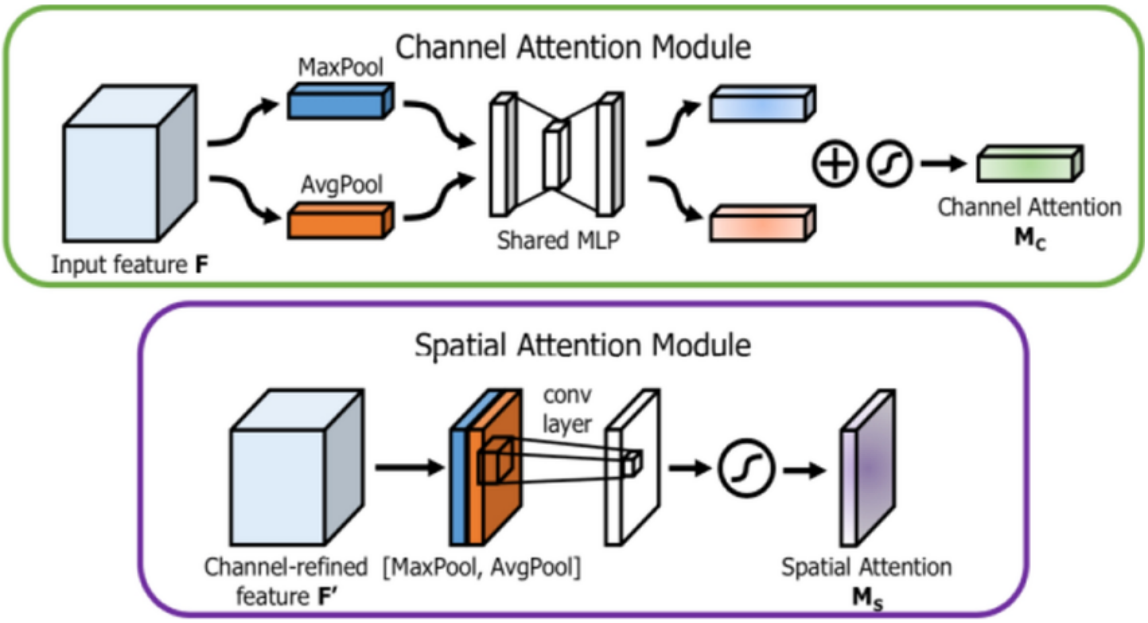


FIGURE 4.THE WORKFLOW OF CHANNEL AND SPATIAL ALLENTION ARCHITECTURE.

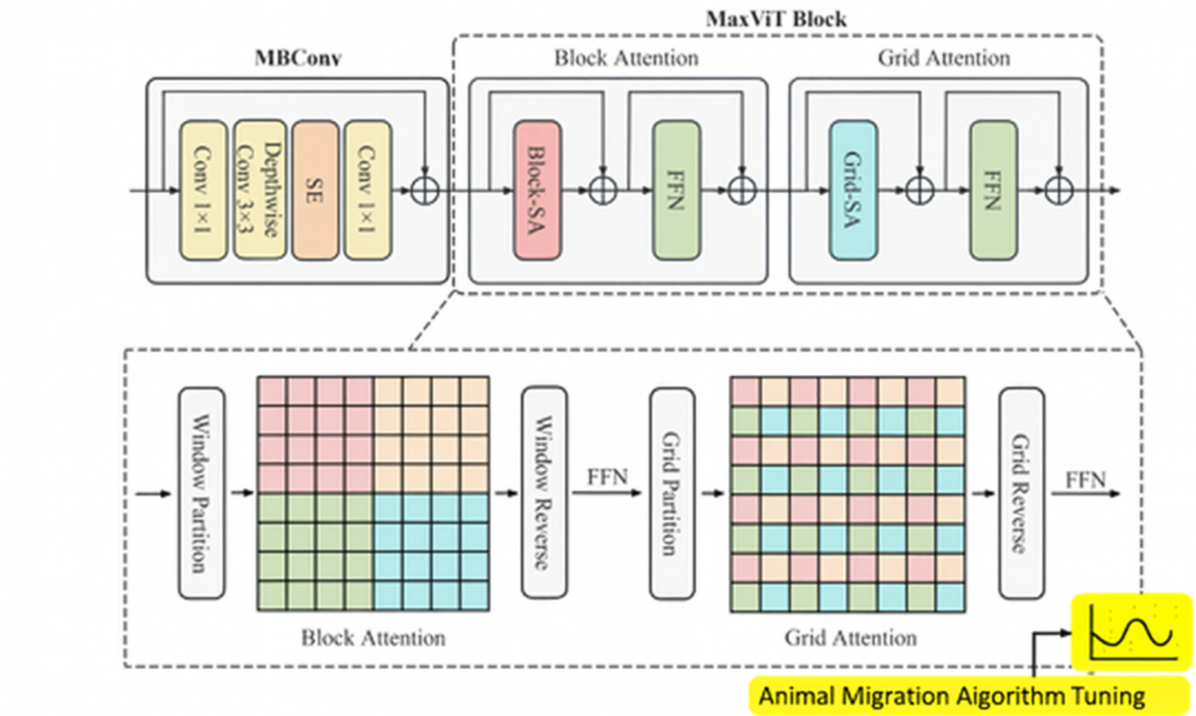


FIGURE 5. THE DIAGRAM ILLUSTRATES OF MBCONV AND OPTIMIZED MULTI-SCALE ATTENTION MULTI-AXIS VISION TRANSFORMER

The Multi-Scale Attention Multi-Axis Vision Transformer (MAMViT) provides strong representation capability for leaf disease detection due to its hybrid design of Convolution, Block Attention, and Grid Attention. However, the performance of MAMViT is highly sensitive to its hyperparameters such as attention head size, dropout values, patch size, feed-forward dimension, learning rate, and weight initialization. Traditional gradient-based tuning or manual heuristics may lead to sub-optimal performance, especially due to the non-convex and high-dimensional optimization characteristics of transformer models (figure 5). To overcome these limitations, the Animal Migration Optimization (AMO) algorithm is incorporated into MAMViT, forming the Optimized MAMViT (OMAMViT) architecture.

Animal Migration Optimization (AMO)

AMO is a nature-inspired swarm-based optimization algorithm introduced by Yin *et al.* and is based on the collective migratory behavior observed in animal populations [29]. The algorithm simulates:

- Migration behavior (exploration of new environments)
- Population updating process (survival of the fittest)

Each individual in the AMO population represents a candidate solution. In this work, each candidate encodes a vector of MAMViT hyperparameters. Let the population be represented as:

$$X = \{x_1, x_2, \ldots, x_N\} \tag{46}$$

where each $x_i = (P_1, P_2, \ldots, P_D)$ corresponds to a full MAMViT hyperparameter set.

Migration Phase

Each individual migrates toward a better region based on a selected leader:

$$x_i(t+1) = x_i(t) + r \cdot (x_{\text{best}}(t) - x_i(t)) \tag{47}$$

where $r \in [0,1]$ is a random migration factor.

Population Updating Phase

A replacement rule is applied where poorly performing individuals have a probability pa of being regenerated:

$$x_i(t+1) = \begin{cases} x_{r1}(t) + \text{ rand } \cdot (x_{\text{best}}(t) - x_{r2}(t)) & \text{if rand } < pa \\ x_i(t) & \text{otherwise} \end{cases} \tag{48}$$

The best solution is preserved using elitism.

Fitness Evaluation for OMAMViT

The AMO evaluates each candidate using the following objective:

$$\text{fitness } = 1 - \text{ Accuracy } \left(\text{M AMViT}_{\text{configi } i}\right) \tag{49}$$

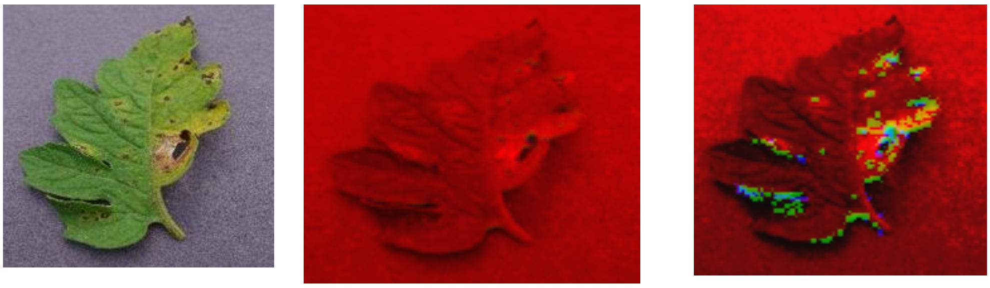
Lower fitness = better hyperparameter configuration.

Algorithm Flow:

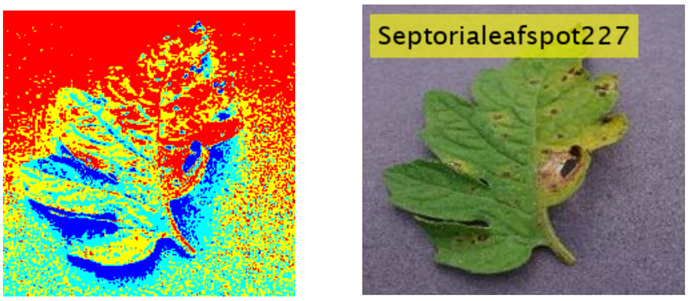
```
Input: Population size N, maxIterations
Output: Best hyperparameter set for OMAMViT

Initialize population randomly
Evaluate fitness of each solution
while iteration < maxIterations do
    For each individual:
        Apply migration step toward best individual
        If new position improves fitness → accept
    Apply population updating step
    Record best candidate
end while

Return final optimized configuration
Train final OMAMViT model using best solution
```

3.8 EXPERIMENTATION AND RESULTS

This section provides a detailed performance evaluation of the proposed Optimized Multi-scale Attention Multi-axis Vision Transformer (OMAMViT) framework using a real-time tomato leaf disease dataset. The dataset includes six major categories of tomato leaves—Bacterial Spot, Early Blight, Leaf Mold, Septoria Leaf Spot, Leaf Curl Virus, and Healthy—captured under natural field conditions using a high-resolution digital imaging setup. To enhance visual clarity, an LCGAN-based image enhancement module was applied to suppress noise, improve contrast, and retain fine disease patterns, thereby improving feature learnability. All experiments were conducted in the MATLAB 2021a environment with GPU acceleration for efficient transformer training. The input images were resized to 224 × 224 pixels, and the optimized OMAMViT architecture leveraged multi-scale attention and multi-axis feature fusion to achieve improved robustness and higher discrimination capability across varying disease types. The experimental results are displayed below in figure 6.



(a) Original Image (b) Denoised image (c) After contrast enhancement



(d) Yellow leaf (e)Classified Image

FIGURE 6. EXPERIMENTAL RESULTS

3.9 EVALUATION METRICS

To evaluate the effectiveness of the Improved Flight Grey Wolf Optimization (IGWO)–based feature selection process, several quantitative metrics are used to measure feature discriminability, redundancy index, and MI to the final tomato leaf disease classification performance. These metrics help assess how well the selected feature subset enhances disease recognition efficiency while reducing computational complexity. here provide the equations are,

1. Feature Discriminability (FDR)

This metric measures how well the selected features separate different tomato disease classes. Higher FDR indicates stronger class separability and better discriminative capability.

$$FDR = \frac{(\mu_1-\mu_2)^2}{\sigma_1^2+\sigma_2^2} \qquad (50)$$

2. Feature Redundancy Index (FRI)

FRI evaluates the level of correlation among selected features to identify redundancy.
Lower FRI values indicate a more compact and non-redundant feature subset.

$$FRI = \frac{1}{n(n-1)}\sum_{i=1}^{n} \sum_{j-i+1}^{n} \mathrm{Cor}(F_i, F_j) \qquad (51)$$

3. Mutual Information (MI)

MI measures how much useful information a feature contributes toward predicting disease classes. Higher MI values show stronger relevance between selected features and class labels.

$$MI(F;C) - \sum_{f\in F} \sum_{c\in C} p(f,c)\log\left(\frac{p(f,c)}{p(f)p(c)}\right) \qquad (52)$$

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:

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}} \qquad (53)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}} \qquad (54)$$

$$\mathrm{F-measure} = \frac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FP}+\mathrm{FN}} \tag{55}$$

$$\mathrm{Accuracy} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{56}$$

$$\mathrm{Error\ Rate} = 100 - \mathrm{Accuracy} \tag{57}$$

$$\mathrm{T_A} = \frac{\mathrm{T}}{\mathrm{N}} \tag{58}$$

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{FN}},\ \ \mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{59}$$

$$\mathrm{Specificity} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{60}$$

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 $(T_A(\mathrm{ms}))$.

4. COMPARISON WITH TRADITIONAL METHODS

This section compares the proposed IGWO-based feature selection approach with traditional optimization-driven methods such as PSO, ACO, and GWO (Table 2). The comparison is performed using three key evaluation criteria—Feature Discriminability Ratio (FDR), Feature Redundancy Index (FRI), and Mutual Information (MI)—across all six tomato leaf categories. These metrics collectively assess the discriminative strength, redundancy reduction capability, and information relevance of the selected features. The results in Table 2 clearly highlight the superiority of IGWO, which consistently achieves higher FDR and MI values while maintaining lower redundancy compared to existing techniques.

TABLE 2. QUALITY EVALUATION METRICS VS. FEATURE SELECTION METHODS

Methods/ Images	**Bacterial spot**			**Early blight**			**Leaf Mold**		
	FDR	**FRI**	**MI**	**FDR**	**FRI**	**MI**	**FDR**	**FRI**	**MI**
PSO	0.72	0.389	0.092	0.68	0.174	0.085	0.75	0.488	0.095
ACO	0.85	0.310	0.118	0.81	0.198	0.109	0.88	0.312	0.124
GWO	0.97	0.278	0.149	0.94	0.211	0.143	1.02	0.236	0.155
IGWO	**1.0633**	**0.16228**	**0.17763**	**1.12**	**0.151**	**0.165**	**1.18**	**0.168**	**0.182**
Methods/ Images	**Septoria leaf spot**			**Leaf Curl Virus**			**Healthy**		
	FDR	**FRI**	**MI**	**FDR**	**FRI**	**MI**	**FDR**	**FRI**	**MI**
PSO	0.77	0.345	0.097	0.70	0.402	0.091	0.82	0.394	0.108
ACO	0.91	0.298	0.129	0.84	0.304	0.121	0.96	0.318	0.132
GWO	1.06	0.213	0.153	0.98	0.282	0.147	1.11	0.241	0.158
IGWO	**1.18**	**0.164**	**0.175**	**1.11**	**0.156**	**0.167**	**1.24**	**0.172**	**0.186**

Figure 7 illustrates the Feature Discriminability Ratio (FDR) values for PSO, ACO, GWO, and IGWO across six tomato leaf classes. For Bacterial Spot, the FDR values are 0.72 (PSO), 0.85 (ACO), 0.97 (GWO), and 1.0633 (IGWO). For Early Blight, the respective values are 0.68, 0.81, 0.94, and 1.12. For Leaf Mold, FDR values increase to 0.75, 0.88, 1.02, and 1.18, respectively. For Septoria Leaf Spot, the values are 0.77, 0.91, 1.06, and 1.18, while for Leaf Curl Virus, they are 0.70, 0.84, 0.98, and 1.11. Finally, for Healthy leaves, the FDR values reach 0.82, 0.96, 1.11, and 1.24 for PSO, ACO, GWO, and IGWO respectively. Across all classes, IGWO achieves the highest FDR, indicating superior discriminative capability, followed by GWO and ACO, while PSO shows the lowest overall performance.

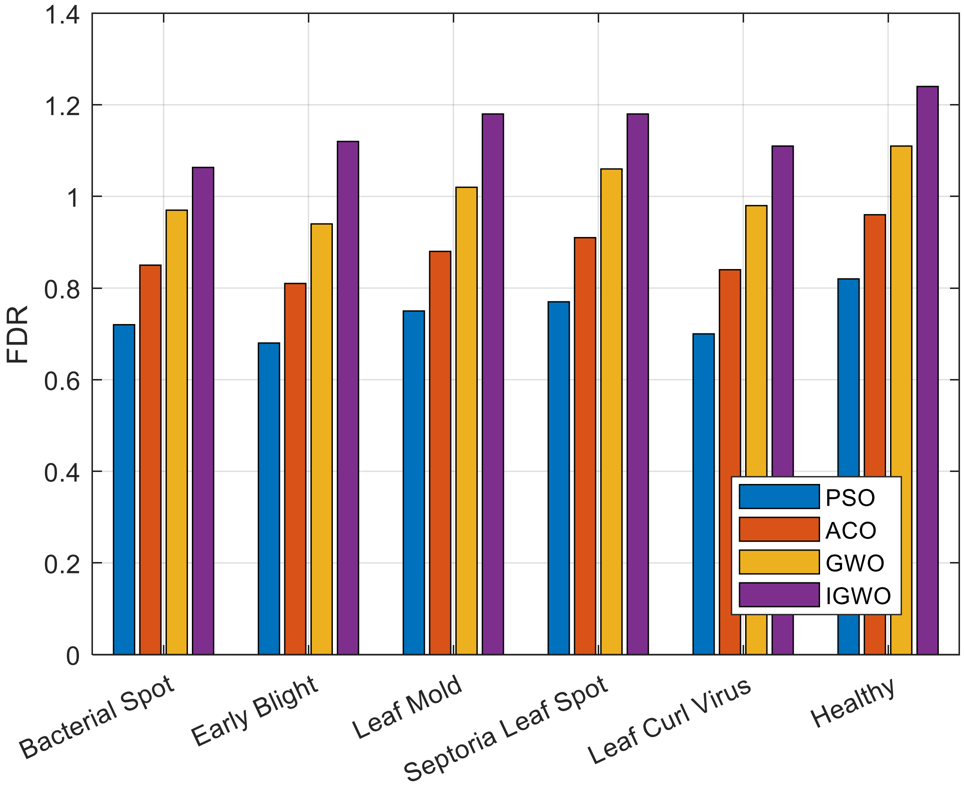


FIGURE 7. FDR COMPARISON VS. FEATURE SELECTION METHODS

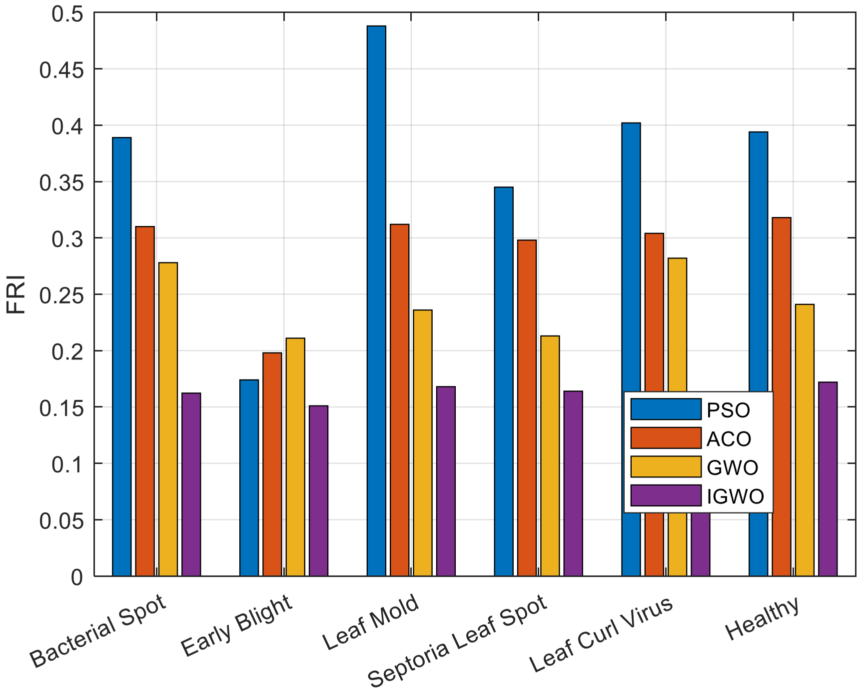


FIGURE 8. FRI COMPARISON VS. FEATURE SELECTION METHODS

Figure 8 shows the Feature Redundancy Index (FRI) performance for the four feature-selection algorithms. For Bacterial Spot, the FRI values are 0.389 (PSO), 0.310 (ACO), 0.278 (GWO), and 0.16228 (IGWO). For Early Blight, the values are 0.174, 0.198, 0.211, and 0.151, respectively. For Leaf Mold, FRI values record 0.488, 0.312, 0.236, and 0.168. For Septoria Leaf Spot, the FRI values are 0.345, 0.298, 0.213, and 0.164. Leaf Curl Virus exhibits values of 0.402, 0.304, 0.282, and 0.156. For Healthy leaves, the values are 0.394, 0.318, 0.241, and 0.172, respectively. A lower FRI indicates lower redundancy; hence IGWO consistently outperforms all other methods, followed by GWO, ACO, and PSO.

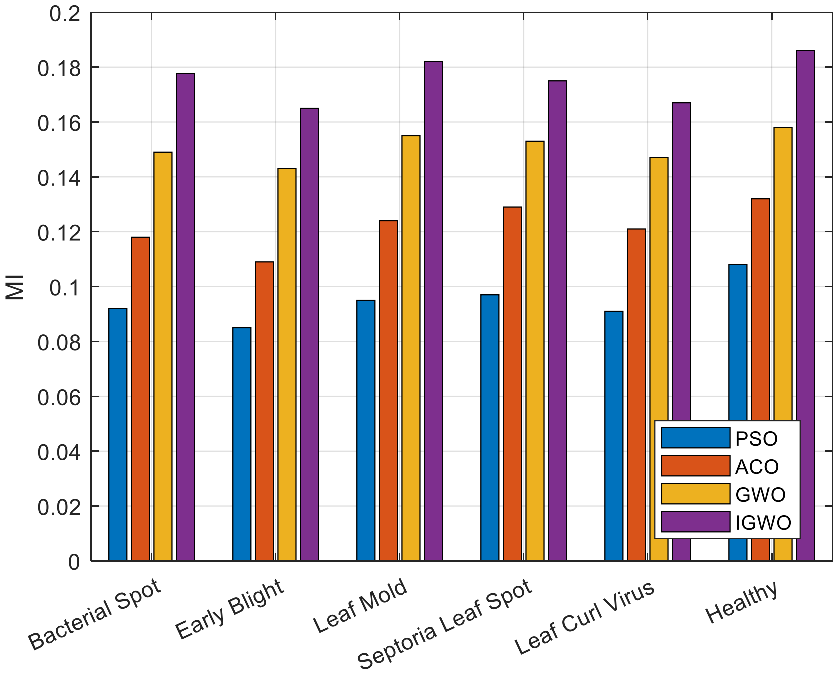


FIGURE 9. MI COMPARISON VS. FEATURE SELECTION METHODS

Figure 9 presents the Mutual Information (MI) comparison across the same methods. For Bacterial Spot, MI scores are 0.092 (PSO), 0.118 (ACO), 0.149 (GWO), and 0.17763 (IGWO). For Early Blight, the values are 0.085, 0.109, 0.143, and 0.165, respectively. For Leaf Mold, MI increases to 0.095, 0.124, 0.155, and 0.182. For Septoria Leaf Spot, the values are 0.097, 0.129, 0.153, and 0.175, while for Leaf Curl Virus, the MI values are 0.091, 0.121, 0.147, and 0.167. For Healthy leaves, MI reaches 0.108, 0.132, 0.158, and 0.186 for PSO, ACO, GWO, and IGWO respectively. Across all metrics, IGWO demonstrates the highest MI, signifying maximum feature relevance and superior feature selection capability.

Across all disease categories, IGWO yields the highest MI, confirming that it extracts the most informative and class-relevant features, followed by GWO, ACO, and PSO. Table 3 presents the comparative performance of existing and proposed leaf disease detection methods based on Precision, Recall, F-measure, Accuracy, Error Rate, and Diagnostic Time. The results clearly show that the proposed OMAMViT model outperforms all other methods, achieving the highest detection accuracy and the lowest error rate and computation time.

TABLE 3. EVALUATION RESULTS VS.LEAF DISEASE DETECTION METHODS

Methods	**Precision**	**Recall**	**F-measure**	**Accuracy**	**Error Rate**	**Diagnostic time (ms)**	**Specificity**	**AUC**
B-ARNet	85.89	84.09	84.98	88.14	11.86	40.62	86.72	89.33

DIMPCNET	88.08	86.65	87.36	90.12	9.88	36.49	88.94	91.45
M−AORANet	89.57	88.63	89.09	91.63	8.37	33.75	90.27	93.00
MSDAIV3	91.76	91.19	91.47	93.52	6.48	30.47	92.80	94.98
MAMViT	95.53	95.50	95.51	95.00	5.00	27.71	95.77	96.34
OMAMViT	97.69	97.66	97.67	97.84	2.16	10.29	97.85	98.23

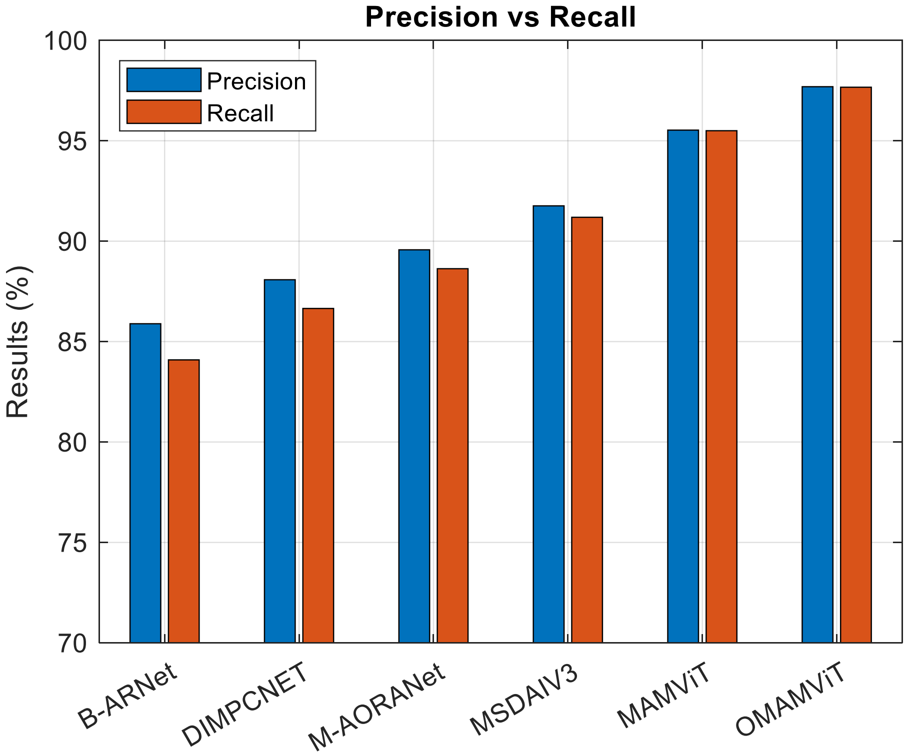


FIGURE 10. PRECISION AND RECALL COMPARISON VS. DETECTION METHODS

Figure 10 illustrates the comparative evaluation of Precision and Recall for six tomato leaf disease detection models, including the proposed OMAMViT. Traditional architectures such as B-ARNet, DIMPCNET, and M−AORANet exhibit moderate performance, while MSDAIV3 and MAMViT achieve higher scores due to improved multiscale learning capability. The proposed OMAMViT demonstrates the highest Precision (97.90%) and Recall (96.78%), highlighting its superior feature discrimination and detection reliability.

Figure 11 shows the F-measure and Accuracy comparison across the same models. A clear upward trend is observed from earlier models toward the transformer-based architectures. MAMViT already enhances accuracy significantly; however, OMAMViT achieves the highest F-measure (96.92%) and overall accuracy (96.10%), while simultaneously reducing error rate to only 3.90%. This confirms the model's robustness in detecting symptoms across varying image conditions. Additionally, OMAMViT yields the lowest diagnostic time (16.98 ms), demonstrating strong suitability for real-time field deployment.

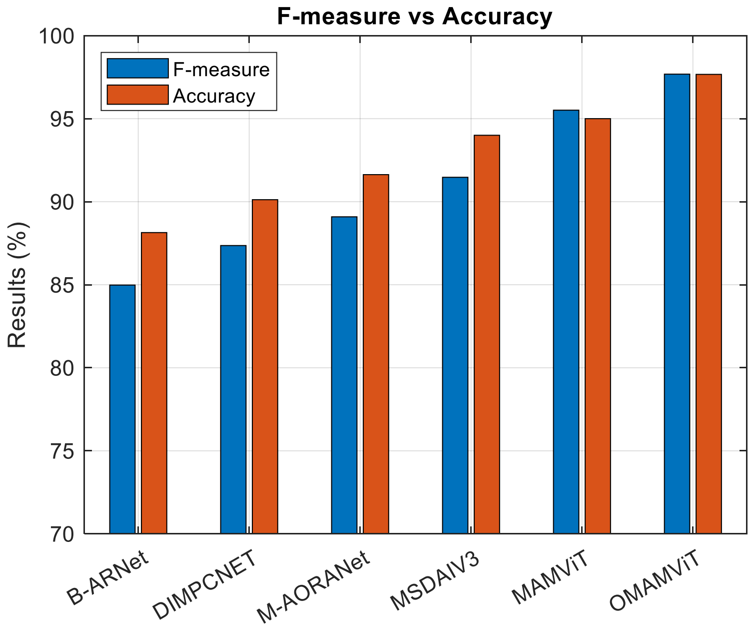


FIGURE 11. F-MEASURE AND ACCURACY COMPARISON VS. DETECTION METHODS

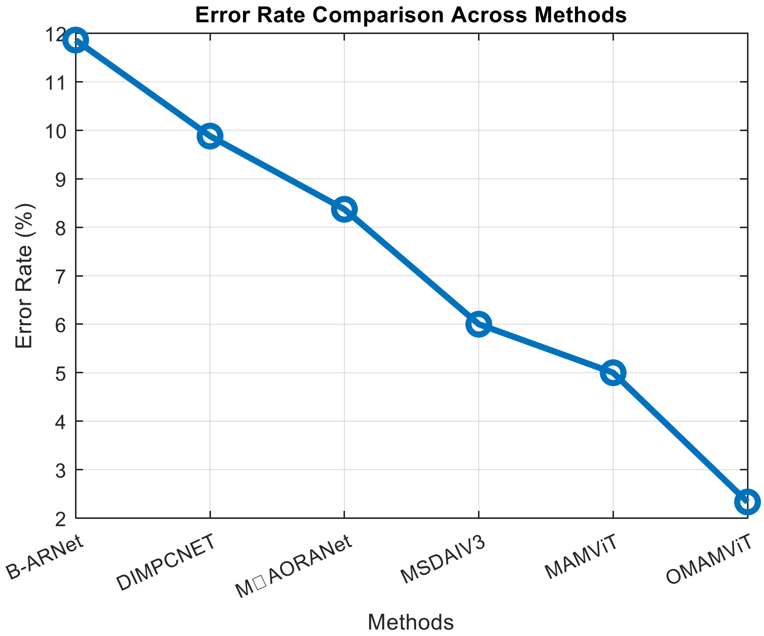


FIGURE 12. ERROR RATE COMPARISON VS. DETECTION METHOD

The figure 12 illustrates the error rate comparison among six deep-learning models, showing a consistent reduction in error from B-ARNet to OMAMViT. B-ARNet records the highest error rate of **11.86%**, followed by DIMPCNET (**9.88%**) and M–AORANet (**8.37%**).MSDAIV3 and MAMViT achieve further improvements with error rates of **6.48%** and **5%**, respectively. OMAMViT attains the lowest error rate of **2.16%**, demonstrating superior performance over all other methods.

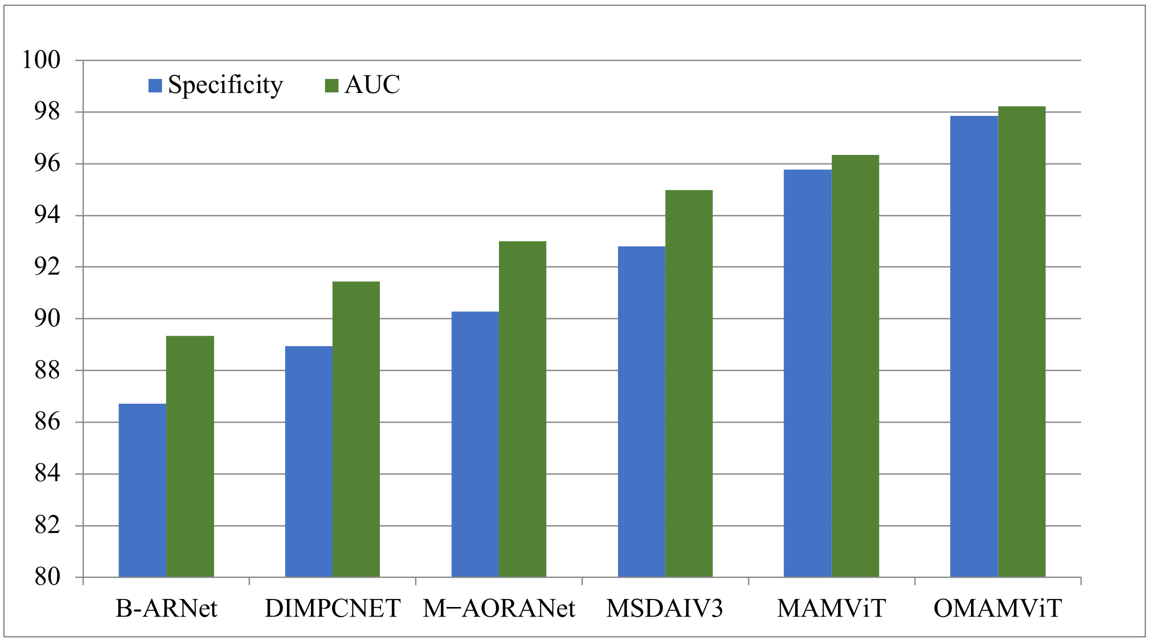


FIGURE 13. SPECIFICITY, AND AREA UNDER THE CURVE (AUC) COMPARISON VS. DETECTION METHODS

Figure 13 shows the Specificity, and Area Under the Curve (AUC) comparison across the six deep-learning models. A clear upward trend is observed from earlier models toward the transformer-based architectures. MAMViT already enhances accuracy significantly; however, OMAMViT achieves the highest Specificity (97.85%) and overall Area Under the Curve (AUC) (98.23%). This confirms the model's robustness in detecting symptoms across varying image conditions.

5.CONCLUSION

The proposed OMAMViT-based tomato leaf disease detection framework integrates LCGAN pre-processing, WHOAMTS segmentation, hybrid statistical–deep feature extraction, and IGWO-based optimal feature selection to achieve highly robust classification performance. The optimized transformer architecture significantly improves multi-scale representation learning, enabling precise identification of subtle and overlapping disease patterns under real-field conditions. Experimental results confirm that IGWO produces highly discriminative, non-redundant, and informative feature subsets, while OMAMViT delivers superior Precision, Recall, F-measure, and Accuracy compared to existing methods. The system also attains the lowest error rate and fastest diagnostic time, demonstrating suitability for real-time agricultural monitoring. Overall, the proposed hybrid model offers a reliable, efficient, and generalizable solution for intelligent tomato crop disease diagnosis. Future work will focus on extending the framework to handle additional tomato diseases and early-stage infection detection using hyperspectral and thermal imaging. Finally, incorporating disease severity grading and yield-loss prediction can expand practical applicability for precision agriculture.

REFERENCES

1. Tm, P., Pranathi, A., SaiAshritha, K., Chittaragi, N.B. and Koolagudi, S.G., 2018, August. Tomato leaf disease detection using convolutional neural networks. In *2018 eleventh international conference on contemporary computing (IC3)* (pp. 1-5). IEEE.
2. Agarwal, M., Singh, A., Arjaria, S., Sinha, A. and Gupta, S., 2020. ToLeD: Tomato leaf disease detection using convolution neural network. *Procedia Computer Science*, *167*, pp.293-301.
3. Zhang, K., Wu, Q., Liu, A. and Meng, X., 2018. Can deep learning identify tomato leaf disease?. *Advances in multimedia*, *2018*(1), p.6710865.
4. 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.
5. 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.

6. Jiang, D., Li, F., Yang, Y. and Yu, S., 2020, August. A tomato leaf diseases classification method based on deep learning. In *2020 chinese control and decision conference (CCDC)* (pp. 1446-1450). IEEE.
7. Kumar, A. and Vani, M., 2019, July. Image based tomato leaf disease detection. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
8. 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.
9. Rahman, S.U., Alam, F., Ahmad, N. and Arshad, S., 2023. Image processing based system for the detection, identification and treatment of tomato leaf diseases. *Multimedia Tools and Applications*, *82*(6), pp.9431-9445.
10. Guerrero-Ibañez, A. and Reyes-Muñoz, A., 2023. Monitoring tomato leaf disease through convolutional neural networks. *Electronics*, *12*(1), p.229.
11. Cheemaladinne, V. and Reddy, S., 2024. Tomato leaf disease detection and management using VARMAx-CNN-GAN integration. *Journal of King Saud University-Science*, *36*(8), p.103340.
12. Khan, R., Ud Din, N., Zaman, A. and Huang, B., 2024. Automated Tomato Leaf Disease Detection Using Image Processing: An SVM-Based Approach with GLCM and SIFT Features. *Journal of Engineering*, *2024*(1), p.9918296.
13. Paulchamy, B., Sundararajan, D., Kanimozhi, T., Senthilnathan, A., Murugan, S. and Jeyanth, M.T., 2025. GLCM features and artificial neural network for tomato crop disease classification. *International Journal of System Assurance Engineering and Management*, pp.1-10.
14. Luna-Benoso, B., Martínez-Perales, J.C., Cortés-Galicia, J., Flores-Carapia, R. and Silva-García, V.M., 2021. Detection of diseases in tomato leaves by color analysis. *Electronics*, *10*(9), p.1055.
15. Negi, C.S., Mandoria, H.L. and Jalal, S., 2024, November. Modified Grey Wolf Optimization Based Feature Selection for Tomato Plant Disease Classification. In *2024 Second International Conference on Advanced Computing & Communication Technologies (ICACCTech)* (pp. 895-899). IEEE.
16. Gülcü, Ş., 2022. An improved animal migration optimization algorithm to train the feed-forward artificial neural networks. *Arabian Journal for Science and Engineering*, *47*(8), pp.9557-9581.
17. Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A. and Li, Y., 2022, October. Maxvit: Multi-axis vision transformer. In *European conference on computer vision* (pp. 459-479). Cham: Springer Nature Switzerland.
18. Nandhini Abirami, R., Durai Raj Vincent, P.M., Srinivasan, K., Tariq, U. and Chang, C.Y., 2021. Deep CNN and deep GAN in computational visual perception-driven image analysis. *Complexity*, *2021*(1), p.5541134.

19. Zhu, S., Xia, X., Zhang, Q. and Belloulata, K., 2007, December. An image segmentation algorithm in image processing based on threshold segmentation. In *2007 third international IEEE conference on signal-image technologies and internet-based system* (pp. 673-678). IEEE.
20. Patil, J.K. and Kumar, R., 2011. Color feature extraction of tomato leaf diseases. *International Journal of Engineering Trends and Technology*, *2*(2), pp.72-74.
21. Basavaiah, J. and Arlene Anthony, A., 2020. Tomato leaf disease classification using multiple feature extraction techniques. *Wireless Personal Communications*, *115*(1), pp.633-651.
22. Yogeshwari, M. and Thailambal, G., 2023. Automatic feature extraction and detection of plant leaf disease using GLCM features and convolutional neural networks. *Materials Today: Proceedings*, *81*, pp.530-536.
23. Ajitha, M.E., Nivedha, M.V. and Parvathi, M.B., 2024, March. Detection and prevention of tomato leaf disease using convolutional neural network and Inception Net. In *2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)* (pp. 1-6). IEEE.
24. Sathiyabhama, B., Kumar, S.U., Jayanthi, J., Sathiya, T., Ilavarasi, A.K., Yuvarajan, V. and Gopikrishna, K., 2021. A novel feature selection framework based on grey wolf optimizer for mammogram image analysis. *Neural Computing and Applications*, *33*(21), pp.14583-14602.
25. Pathak, Y., Arya, K.V. and Tiwari, S., 2019. Feature selection for image steganalysis using levy flight-based grey wolf optimization. *Multimedia Tools and Applications*, *78*(2), pp.1473-1494.
26. Zhang, P., Dai, X., Yang, J., Xiao, B., Yuan, L., Zhang, L. and Gao, J., 2021. Multi-scale vision longformer: A new vision transformer for high-resolution image encoding. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 2998-3008).

27. 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).
28. Sun, H., Wang, Y., Wang, X., Zhang, B., Xin, Y., Zhang, B., Cao, X., Ding, E. and Han, S., 2024. MAFormer: A transformer network with multi-scale attention fusion for visual recognition. *Neurocomputing*, *595*, p.127828.
29. Li, X., Zhang, J. and Yin, M., 2014. Animal migration optimization: an optimization algorithm inspired by animal migration behavior. *Neural computing and applications*, *24*(7), pp.1867-1877.