Journal of Theoretical and Applied Information Technology
28th February 2026. Vol.104. No.4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# GRASSHOPPER OPTIMIZED VISION TRANSFORMER WITH ADAPTIVE TOKEN CALIBRATION DROPOUT SCALING QKV ATTENTION REFINEMENT IN COTTON LEAF DISEASE IDENTIFICATION

**K. KARTHIGA[1], Dr. B. RAJDEEPA[2]**

[1]Research Scholar, Department of Computer Science, PSG College Of Arts & Science, Coimbatore, India
[2] Associate Professor and Head, Department of Information Technology, PSG College Of Arts & Science, Coimbatore, India
E-mail: [1]karthigakanthanm@gmail.com, [2]rajdeepa@psgcas.ac.in

**ABSTRACT**

Accurate identification of cotton leaf diseases remains a complex task due to high visual similarity among disease types, irregular lesion boundaries, background interference, and the early-stage subtlety of symptoms in field-captured images. Current literature lacks transformer-based disease identification frameworks that dynamically regulate attention, token relevance, and learning stability in response to lesion morphology. This study bridges that gap by introducing an optimization-driven attention control mechanism that generates new insights into lesion-focused transformer learning for precision agriculture. Traditional models and vanilla Vision Transformers struggle with fixed patch segmentation, redundant token processing, and diluted attention focus, leading to suboptimal classification. To address these challenges, this research proposes a novel model titled Grasshopper Optimized Vision Transformer (GO-ViT), engineered for lesion-aware feature refinement in cotton leaf disease identification. GO-ViT integrates swarm-based Grasshopper Optimization to dynamically configure patch embedding size, attention head scaling, dropout distribution, QKV matrix refinement, and entropy-based token pruning. This biologically inspired regulation emulates adaptive foraging patterns, ensuring high lesion focus and reduced attention waste. GO-ViT enhances inter-class separation, preserves spatial continuity, and minimizes false activation over non-symptomatic regions. Experimental evaluation across five performance metrics confirms GO-ViT's robustness, with 94.491% Overall Detection Efficiency and 88.985% Balanced Class Correlation. The model demonstrates superior ability to isolate overlapping infections, suppress background noise, and deliver stable predictions across diverse cotton leaf conditions. GO-ViT stands as a reliable solution for scalable, high-precision disease classification in precision agriculture, combining visual attention learning with nature-driven optimization control.

**Keywords:** *Cotton Leaf Disease, Vision Transformer, Grasshopper Optimization, Token Pruning, Attention Calibration, Lesion Segmentation*

## 1. INTRODUCTION

Cotton remains one of the most economically important crops cultivated across several agroclimatic zones. Its productivity is heavily impacted by foliar diseases, which manifest as bacterial spots, viral mosaics, fungal blights, and nutrient-related discolorations [1]. These infections often emerge progressively over the crop's growth cycle, initially displaying subtle symptoms that are difficult to detect without specialized equipment. Traditional inspection methods suffer from observer bias and inconsistency, especially under fluctuating light or partial occlusions on field-grown leaves [2]. Visual similarities among diseases further complicate accurate field diagnosis. Moreover, early symptoms can mimic those of unrelated stressors, making precise categorization a demanding task. Manual surveillance also lacks the scalability to handle widespread acreage, resulting in delayed treatment decisions. Such limitations call for more structured, data-driven identification systems that capture temporal variations and provide consistent, stage-independent detection of cotton leaf disorders [3], [4].

Image-based approaches have become the preferred route for automating disease identification [5]. Field-acquired leaf images often capture both symptomatic and non-symptomatic zones, as well as natural noise elements such as dirt, overlapping leaves, and uneven lighting. These visual inconsistencies affect the performance of conventional models, which operate on fixed-resolution features [6], [7]. Lesion boundaries in cotton leaves are irregular and may

span multiple small image patches. When processed individually, these patches can fragment lesion structure, leading to false negatives or ambiguous predictions. This scenario requires systems that preserve lesion continuity, dynamically assess patch relevance, and adapt to fine-grain and wide-area symptoms. Disease categories frequently overlap in texture and color tone, increasing intra-class ambiguity and necessitating high discriminative sensitivity. Addressing these challenges requires a model architecture that can capture hierarchical spatial patterns while remaining resilient to visual distortions and patch-level inconsistencies [8].

Transformer-based architectures, renowned for their global attention capabilities, provide an avenue for rethinking image classification in disease diagnostics. These models divide images into patches, embed them into token sequences, and use self-attention to compute cross-patch dependencies [9]. Unlike convolutional models, transformers do not rely on locality alone and instead learn holistic patterns across the entire image. This ability proves valuable in detecting dispersed or scattered disease zones. Standard Vision Transformers (ViT) suffer from rigid patch processing, excessive memory use, and difficulty focusing on small, biologically essential lesions [10]. Uniform attention allocation often favors dominant backgrounds, while minor lesions that demand higher clinical attention may be underrepresented. The inability to adaptively scale attention or adjust learning priorities during training limits the practical utility of such architectures in noisy, variable agricultural settings [11].

In response, hybrid computational strategies are being explored where attention-based models are enhanced through biologically inspired optimization techniques [12]. These strategies mimic the collective intelligence of species like grasshoppers, where group-based foraging dynamics spanning repulsion, alignment, and attraction guide search behaviors in dynamic environments [13]. When translated into deep learning pipelines, such behaviour models help refine architectural decisions, dynamically adjust learning rates, and regulate feature retention or dropout in response to spatial uncertainty [14]. Grasshopper-inspired optimization supports adaptive control over token selection, attention sharpness, and patch fusion. This aligns closely with the demands for cotton leaf disease identification, where symptoms vary in density, orientation, and visibility across leaf surfaces. By embedding such optimization layers within transformer blocks, the model evolves with precision and generalization, capable of capturing rare disease instances while retaining clarity on dominant ones.

Model scalability and interpretability remain critical for real-world deployment. Disease recognition systems must maintain high classification accuracy and function efficiently under limited computing constraints [15]. This is essential for in-field deployment, where on-device processing is standard. Token pruning and attention-guided memory reduction help lower model complexity without sacrificing lesion-specific focus. At the same time, explainable outputs such as patch-wise confidence maps and lesion heat signatures improve trust among domain users [16]. Such outputs enable agronomists and field agents to correlate model decisions with visible evidence, thereby reinforcing confidence in the system's diagnostic accuracy [17]. These converging ideas transformer-based learning, swarm-guided optimization, and domain-aware adaptability collectively enable a new generation of plant disease detection frameworks tailored for the complexity of agricultural ecosystems [18].

This study concentrates exclusively on image-based identification of cotton leaf diseases using transformer-driven visual learning under field-acquired RGB imagery. The work focuses on spatial attention refinement, token relevance control, and lesion-centric feature representation within a Vision Transformer framework enhanced through grasshopper-inspired optimization. The scope excludes spectral, hyperspectral, thermal, or multisensor fusion data, and does not address disease severity grading, yield loss estimation, or temporal disease progression analysis. The proposed framework is evaluated under controlled experimental datasets and does not consider real-time UAV streaming, cross-crop generalization, or multimodal agronomic inputs. Model performance is assessed under fixed training–testing partitions, assuming reliable disease annotation and consistent image quality. Limitations include dependency on labeled data availability, sensitivity to extreme illumination distortions, and absence of on-device latency benchmarking. These boundaries define the operational context and applicability of the proposed approach.

### 1.1. Problem Statement

Identifying cotton leaf diseases remains a persistent challenge in agricultural diagnosis due to visually similar symptoms that evolve and vary across different stages of infection. Disease regions often span irregular shapes and appear fragmented, making accurate spatial mapping a complex task. Many models fail to maintain lesion continuity across tokenized patches, resulting in fragmented

interpretations of widespread disease patterns. Small lesions with weak contrast are frequently underrepresented, while visually dominant backgrounds distort classification accuracy. The high visual overlap between different diseases increases inter-class confusion, resulting in misdiagnosis. Token redundancy in deep transformer architectures further inflates computational load, limiting their application in real-time or low-resource environments. Attention spread across irrelevant patches, and a lack of adaptive learning prioritization toward minority disease classes weakens diagnostic robustness. These limitations hinder the timely identification of diseases, impacting decision-making and agronomic planning in large-scale cotton cultivation systems.

Existing cotton leaf disease identification studies predominantly rely on convolution-centric pipelines, ensemble voting strategies, or static transformer attention schemes that treat spatial tokens with uniform importance. Literature demonstrates consistent improvements in accuracy, yet reveals persistent shortcomings in lesion continuity preservation, background suppression, and adaptive discrimination under visually overlapping disease categories. Most transformer-based approaches apply fixed patch segmentation and uniform self-attention, which fragments irregular lesion structures and weakens sensitivity to early-stage or low-contrast infections. Optimization techniques reported in prior studies are largely confined to preprocessing, feature selection, or external parameter tuning, leaving internal attention dynamics and token behavior unregulated. This creates a methodological gap where attention diffusion, token redundancy, and class imbalance remain structurally unaddressed. The present study is required to bridge this gap by embedding biologically inspired optimization directly into transformer attention, token pruning, and learning regulation, enabling lesion-driven visual reasoning that current state-of-the-art solutions do not explicitly support.

### 1.2. Motivation

Achieving consistent, high-yield cotton production depends on early and accurate detection of foliar diseases that threaten plant vitality and fiber quality. A single undetected infection can rapidly propagate, reducing photosynthetic potential and introducing yield penalties. Manual scouting methods are time-consuming and inconsistent, particularly across vast cotton fields that are exposed to climatic variability and multi-pathogen risks. Automated diagnostic systems must be able to classify diseases and adapt to diverse environments, including fields with poor lighting or dense vegetation. Precision in detection reduces unnecessary pesticide use, preserves crop health, and improves input efficiency. The motivation arises from building a reliable and scalable system that identifies all disease types—mild or severe, common or rare—across different stages. Field adaptability, lightweight deployment, and the ability to provide timely and interpretable results are essential for supporting real-world cultivation practices and sustainable yield optimization.

### 1.3. Objective

This research aims to develop a novel model named Grasshopper Optimized Vision Transformer (GO-ViT) to address the challenges outlined in the problem statement. GO-ViT enhances disease classification by integrating biologically inspired Grasshopper Optimization into the Vision Transformer architecture. The objective is to refine patch-level feature selection, dynamically adjust attention distributions, and optimize learning rate schedules to enhance training stability. GO-ViT is intended to reduce redundancy through token pruning, refine Q-K-V scaling for lesion saliency, and apply adaptive loss weighting to handle class imbalance. It seeks to improve spatial continuity in lesion representation and mitigate overfitting to dominant backgrounds. GO-ViT also targets computational adaptability, ensuring efficient operation in resource-constrained settings. The model aims to provide a robust, interpretable, and domain-aware solution that can precisely identify cotton leaf disease in both experimental and field-level conditions.

The core aim of this study lies in introducing a lesion-aware transformer framework that departs from static architectural assumptions through adaptive, optimization-driven control of attention, token relevance, and regularization behavior. Novelty is established through the unified integration of grasshopper optimization directly into transformer internals rather than peripheral hyperparameter tuning. Outcome measures extend beyond accuracy to include balanced correlation, prediction deviation, sensitivity–specificity equilibrium, and precision stability under inter-class visual overlap. The study contributes new knowledge by demonstrating that swarm-guided regulation within transformer blocks can systematically align attention flow with lesion

topology, improving robustness under real-field visual variability.

### 1.4. Organization of the Paper

The Introduction focuses on the complexity of detecting cotton foliar diseases under irregular lesion boundaries and spatial noise. The Literature Review presents a general overview of leaf disease detection systems and their limitations in addressing visual ambiguity and large-scale agricultural applications. The Methodology proposes GRASSHOPPER OPTIMIZED VISION TRANSFORMER, a transformer enhanced with grasshopper optimization to refine attention focus, control token pruning, and regulate spatial encoding. The Dataset and Performance Metrics section summarises the nature of collected image samples, class definitions, and evaluation criteria. The Results and Discussion section examines performance in terms of lesion coverage, attention sharpness, and resource efficiency. The Conclusion confirms GO-ViT's strength in maintaining interpretability and spatial precision under field imaging constraints.

## 2. LITERATURE REVIEW

"Multimodal Fusion Framework (MFF)" [19] integrates an image-driven CNN with a time-aware RNN to identify cotton leaf diseases using visual and sequential data. CNN focuses on extracting spatial clues from leaf patterns, while RNN decodes time-based environmental fluctuations such as humidity or temperature shifts. The model gains a balanced view across inputs by fusing decisions through weighted voting. Each stream reinforces the other, capturing surface-level symptoms and deeper contextual trends. The outcome is a cohesive prediction mechanism that adapts to appearance and evolving crop conditions. This framework uses lightweight, interpretable network fusion to promote consistent detection under field uncertainties. "SmartCotton TL-Scan (SCTLS)" [20] deploys fine-tuned transfer learning models for diagnosing cotton leaf diseases. VGG-16, VGG-19, Inception-V3, and Xception are modified by adjusting trainable layers and tuning key parameters. Xception offers optimal results due to its depthwise separable filters. A web-based tool integrates this model to process uploaded leaf images, delivering real-time predictions. The system ensures fast inference, compatibility with low resources, and high accuracy. The model performs field-level image diagnostics with automated resizing, normalization, and classification. This combination facilitates rapid detection, enabling an early response to disease. It is scalable for other crops by retraining within the same architecture-driven interface.

"Reservoir-Light AppleID" [21] utilizes a reservoir computing model with optoelectronic time-delay dynamics for the rapid detection of apple leaf disease. Leaf images are converted into pixel sequences that modulate light intensity and are then processed through a photonic delay system. Only the readout layer is trained using ridge regression. This approach captures high-dimensional temporal lesion patterns, enabling low-latency, energy-efficient computation. Optical encoding preserves spatial features, thereby reducing the need for deep stacks. Outputs are mapped to disease classes, supporting ultra-fast, neuromorphic photonic inference. "Tensor-CNN Tomato Insight" [22] integrates tensor subspace learning with a pre-trained CNN for tomato leaf disease analysis. CNN-extracted features are reshaped into high-order tensors, maintaining spatial and channel interdependencies. Tensor decomposition reduces dimensionality, retaining geometric locality and class separability. Orthogonal projections extract dominant latent modes, suppressing noise. A support vector classifier predicts disease from compressed embeddings. This method avoids redundancy, increases interpretability, and enables accurate detection of subtle disease traits, remaining robust under scale variation and occlusion. "FusionStream Fruit Decoder" [23] implements a two-stream deep learning pipeline for fruit disease classification. One stream extracts holistic features from the whole fruit using a CNN, while the other focuses on lesion-prone areas via region proposals. Channel attention amplifies salient features in both streams. Outputs are fused in a shared latent space using weighted correlation scores. The final softmax classifier categorizes fruit diseases into distinct classes. This dual-branch design strikes a balance between coarse identification and fine-grained analysis, outperforming single-path models in cluttered environments.

"SUNet BrewWatch" [24] introduces a hybrid model fusing spatial and sequential features for coffee leaf disease detection. A convolutional encoder extracts multi-scale descriptors, which a bidirectional LSTM processes to model pixel-level sequence dependencies. A saliency-aware module enhances focus on symptomatic zones. The decoder incorporates skip connections and dilated convolutions to preserve edges. Features are fused and reweighted before classification. The model utilizes focal loss to address class imbalance,

enabling robust localization even in the presence of overlapping foliage and low contrast. "StageWise Wheat Ensemble" [25] deploys a multi-stage ensemble framework for wheat disease detection. Each stage uses an independent CNN with different receptive fields, refining texture features from coarse to fine. Probabilistic outputs feed into subsequent stages, integrating edge detectors and wavelet features. Outputs are fused using gated attention, and a final ensemble layer conducts weighted majority voting. Regularization reduces overfitting. The modular architecture supports effective learning with limited samples and can be transferred to other cereal crops. "AshEye UAV Monitor" [26] uses UAVs to capture high-resolution leaf images for real-time ash dieback detection and severity quantification. A segmentation network isolates leaves, and a CNN detects necrotic, chlorotic, and scorched regions. Severity indices are calculated based on pixel intensity and lesion coverage. A spatial attention layer maps disease clusters, estimating the spread of the canopy. UAV flight paths are optimized dynamically. The system updates severity metrics, transmits alerts, and integrates GPS for targeted surveillance and precision intervention.

"YOLOv7 Tomato Scanner" [27] combines object detection and classification for the diagnosis of tomato diseases. A modified YOLOv7 detects diseased leaf regions with adaptive anchor resizing and attention-integrated feature pyramids. Detected patches undergo multi-label classification via a CNN. Spatial attention refines feature maps, distinguishing overlapping symptoms. Batch normalization and residual connections stabilize training. Output layers support simultaneous multi-disease prediction, synchronizing bounding box confidence, and labels for spatial mapping. This architecture enables accurate localization and classification in real-world conditions. "PCA-Deep Tomato Classifier" [28] integrates PCA with a deep neural network for tomato leaf disease classification. PCA reduces redundant spatial dimensions, preserving disease-related variance. Transformed inputs feed into a hierarchical CNN with asymmetric kernels for fine lesion detection. Layer-wise feature compression prevents overfitting. Intermediate residual blocks preserve transitional features. The classifier uses softmax with cosine distance loss. Augmentation with Gaussian noise and rotation mimics farm variability. Filter pruning guided by PCA improves feature extraction efficiency for agro-processing. "ExplainLeaf Ensemble" [29] unifies ensemble deep models with interpretable layers for plant disease identification. Three parallel CNNs, each trained on distinct augmentations, capture view-invariant symptoms. Outputs are blended using weighted soft voting. LIME-based explainability layers generate heatmaps for model decisions, aiding agronomist interpretation. Base learners include DenseNet, EfficientNet, and ResNet, which are optimized with focal loss for addressing class imbalance. Confidence thresholds trigger retraining on ambiguous cases. This ensemble delivers accurate predictions and clear visual explanations for disease classification.

The "MetaFusion Disease Engine" [12] constructs a feature-ensemble system that integrates CNN embeddings and metaheuristic fusion. Three CNNs extract diverse representations, concatenated and reduced via t-SNE. Metaheuristic algorithms (such as Grey Wolf, Particle Swarm, and Ant Colony) rank feature subsets based on their classification relevance. SVM evaluates each subset, selecting the best for the final prediction. Comparative analysis ensures robustness and optimal feature consolidation. Adaptive weighting and hybrid learning strike a balance between computational cost and diagnostic precision, supporting robust plant disease detection across diverse datasets. "PlantDet FusionCore" [30] deploys a structured ensemble of pre-trained CNNs fused by adaptive weighted voting for plant disease detection. The framework includes preprocessing, multi-model feature extraction, and decision-level fusion. VGG19, ResNet50, and InceptionV3 generate diverse representations, each trained on bootstrapped samples. Model weights are tuned on a validation set. The fusion module integrates decisions with dynamic weighting based on symptom complexity. A confidence layer flags uncertain cases for manual review, ensuring consistent predictions across various leaf types and conditions. "CA-Lite Diagnoser" [31] leverages an optimized convolutional structure with channel attention and refined up-sampling for crop disease detection. A compact CNN backbone extracts shallow features, with channel attention prioritizing lesion-informative filters. Lightweight up-sampling using pixel shuffle patterns refines spatial resolution efficiently. Multi-scale convolution captures low and high-frequency lesion textures. Outputs from different levels are fused using attention-weighted summation. This design compresses the model without sacrificing classification precision, supporting accurate, efficient disease prediction.

The increasing complexity and ambiguity in visual disease patterns demand decision

frameworks that can adapt under uncertainty rather than rely on rigid feature hierarchies. Bio-inspired optimization offers an effective mechanism to emulate adaptive search and selective focus observed in natural systems, enabling dynamic regulation of attention, relevance, and convergence within deep learning models [32]-[71]. Plithogenic Cubic Sets complement this paradigm by formally modeling indeterminacy, contradiction, and partial membership that arise when lesion characteristics overlap across disease classes. Integrating bio-inspired optimization with Plithogenic Cubic Sets provides a principled foundation for handling conflicting visual evidence, uncertain boundary regions, and graded lesion presence, which are inadequately addressed in conventional crisp or fuzzy representations [72]-[74]. This alignment supports adaptive learning under uncertainty while preserving structured decision interpretability in complex agricultural disease identification tasks.

The reviewed literature highlights strong advancements in convolutional learning, ensemble fusion, and hybrid architectures for plant disease detection, yet reveals limited exploration of adaptive attention governance within transformer-based vision models. Existing approaches emphasize feature extraction or decision-level fusion without directly addressing attention diffusion, token redundancy, or lesion fragmentation introduced by fixed patch encoding. Optimization techniques in prior work remain external to the learning pipeline, functioning primarily as parameter selectors rather than structural regulators. This study responds to this unmet need by embedding biologically inspired optimization within the attention and token dynamics themselves, establishing a closer alignment between model behavior and lesion morphology.

## 3. GRASSHOPPER OPTIMIZED VISION TRANSFORMER (GO-ViT)

The proposed methodology introduces an integrated diagnostic flow under the framework named Grasshopper Optimized Vision Transformer (GO-ViT), designed to overcome the limitations in cotton leaf disease identification. Attention-based modules are configured with spatial sensitivity, while optimization layers adaptively regulate learning dynamics, feature emphasis, and computational efficiency. The design rejects fixed configurations and instead adjusts to lesion irregularities, patch-level ambiguity, and category-wise imbalance. Each core function—from patch merging to dropout control—is developed to handle real-field image complexity where infection zones appear with varied textures and unclear borders. The structured progression reflects how GO-ViT responds to uncertainty in agricultural imaging.

### 3.1. GO-Based Hyperparameter Initialization

In the behavioral ecology of grasshoppers, swarm patterns emerge through a delicate balance of attraction and repulsion forces among individuals. This natural mechanism underpins the search efficiency of GO introduced in high-dimensional spaces. Within the context of hyperparameter initialization for the Vision Transformer (ViT), GO mimics this intelligent group navigation to explore various configuration spaces. Patch size, attention headcount, dropout rate, and L2 regularization strength significantly influence the capacity of ViT to retain critical spatial cues from cotton leaf imagery. Random initialization often results in inconsistent convergence or essential loss of lesion characteristics. GO-based search allows coverage of a diverse population of initialization vectors, steering toward promising regions of the configuration space while preserving heterogeneity among candidates. Let $H_i$ denote the $i^{th}$ hyperparameter vector in the search space. The position update function integrates social interaction, environmental cues, and search direction encoded in the GO mechanism, and each hyperparameter update is performed using Eq. (1), controlled by three separate interaction dynamics. The attraction factor modulates convergence toward fitter configurations, the direction term introduces preference for elite solutions based on current evaluation, and the perturbation ensures the search avoids entrapment in locally optimal configurations.

$$H_i^{(t+1)} = H_i^{(t)} + A_i^{(t)} \cdot D_i^{(t)} + R_i^{(t)} \cdot \zeta^{(t)} \qquad (1)$$

Where, $A_i^{(t)}$signifies the attraction factor guiding exploration, $D_i^{(t)}$ models directional bias influenced by global best configurations, and $R_i^{(t)} \cdot \zeta^{(t)}$ introduces adaptive perturbations for diversity enhancement.

Each position vector of a grasshopper encodes a specific ViT setup, transforming the biological abstraction into a structured set of model parameters. Patch size $(p_s)$, attention head count $(a_h)$, dropout probability $(d_p)$, and L2 weight decay $(w_r)$ are the primary variables under optimization. These variables are normalized to the range [0, 1] and mapped to discrete or continuous model settings. The mapping function $\Phi(H_i^{(t)})$ transforms the

normalized position vector to actual hyperparameter values:

$$\Phi\left(H_i^{(t)}\right) = \left[\left\lfloor s_{min} + (s_{max} - s_{min}) \cdot H_{i1}^{(t)} \right\rfloor, \left\lfloor h_{min} + (h_{max} - h_{min}) \cdot H_{i2}^{(t)} \right\rfloor, d_{min} + (d_{max} - d_{min}) \cdot H_{i3}^{(t)}, r_{min} + (r_{max} - r_{min}) \cdot H_{i4}^{(t)}\right] \quad (2)$$

Where, $H_{i1}^{(t)}$ through $H_{i4}^{(t)}$ denote the components of the $i^{th}$ position vector and the respective ranges are defined for each parameter. Patch sizes and attention heads undergo flooring to ensure discrete model instantiations, while dropout and regularization maintain their continuous nature for gradient-based fine-tuning.

The interaction between grasshoppers in swarm behavior is fundamentally defined by relative distances and directionality, translating to hyperparameter contrast and performance differential. Eq. (3) ensures that grasshoppers distant in parameter space contribute less to positional adjustment, thereby maintaining a swarm-wide balance between convergence toward high-performance regions and dispersion for coverage.

$$S_{ij}^{(t)} = \beta \cdot \frac{H_j^{(t)} - H_i^{(t)}}{\left\| H_j^{(t)} - H_i^{(t)} \right\|} \cdot e^{-\gamma \cdot \left\| H_j^{(t)} - H_i^{(t)} \right\|} \quad (3)$$

The term $S_{ij}^{(t)}$ captures the social force from peer $j$ on individual $i$, weighted by a distance-dependent exponential decay. $\beta$ controls the magnitude of attraction, while $\gamma$ modulates repulsion based on configuration dissimilarity. The exponential decay ensures proximity-induced influence and minimizes redundant exploration.

A bounding mechanism is applied post-update to maintain biological plausibility while constraining parameter evolution within feasible regions. Let $\Theta\left(H_i^{(t+1)}\right)$ denote the bounded transformation:

$$\Theta\left(H_i^{(t+1)}\right) = \begin{cases} 1 & if \quad H_{ik}^{(t+1)} > 1 \\ 0 & if \quad H_{ik}^{(t+1)} < 0 \\ H_{ik}^{(t+1)} & otherwise \end{cases} \quad (4)$$

This threshold-based bounding scheme is applied to all components of the hyperparameter vector. Any excess over the range is trimmed, ensuring meaningful mapping during decoding into the ViT configuration. The key objective remains maximizing classification accuracy while retaining disease-specific pattern sensitivity in the cotton leaf dataset. A dynamic weighting mechanism enhances parameter updates by incorporating classification error gradients. The new directional weighting for each agent is expressed as:

$$\Delta H_i^{(t)} = \eta \cdot \frac{\partial L\left(H_i^{(t)}\right)}{\partial H_i^{(t)}} \quad (5)$$

Where, $\Delta H_i^{(t)}$ adjusts each hyperparameter component based on the loss gradient $\partial L / \partial H_i$, scaled by learning rate $\eta$. The loss function $L$ is constructed to penalize misclassification across underrepresented disease categories, enhancing sensitivity to rare infections.

As classification progresses, the optimizer adapts its population diversity based on the deviation in attention maps across different ViT instances. Let $\mu_A$ and $\sigma_a$ represent the mean and standard deviation of self-attention scores across all heads in each configuration. Population diversity adjustment follows:

$$k^{(t)} = \frac{1}{1 + e^{-\delta\left(\sigma_A^{(t)} - \mu_A^{(t)}\right)}} \quad (6)$$

The diversity controller $k^{(t)}$ acts as a sigmoid function of attention variation, with $\delta$ regulating sensitivity. This dynamic population resizing prevents premature saturation in the configuration space and maintains optimal swarm responsiveness.

### 3.2. Dynamic Patch Merging for Lesion Continuity

ViT divides an image into non-overlapping square patches. Such discretization frequently disrupts visual coherence, especially in images containing organic lesion structures typical of cotton leaf diseases. Standard patch segmentation may fragment key lesion patterns in scenarios where disease spreads across adjacent patches, resulting in distorted self-attention maps and reduced model sensitivity. This fragmentation affects disease classification by diffusing discriminative lesion cues across disjoint tokens. GO, inspired by the clustering dynamics observed in swarming, adapts local exploration strategies to merge spatially adjacent patches with correlated disease intensity. The dynamic merging process prioritizes patches with overlapping activation zones corresponding to infected tissue, ensuring lesion continuity through

spatial aggregation. The adjacency intensity map $I_{i,j}$ between two patches $i$ and $j$ can be expressed as:

$$I_{i,j} = \frac{1}{Z}\sum_{k=1}^{K}\left|P_i^k - P_j^k\right| \quad (7)$$

Where, $P_i^k$ and $P_j^k$ represent the pixel intensity values for the $k^{th}$ element of patch $i$ and patch $j$, and $Z$ is a normalization constant. This formulation captures local lesion similarity necessary for merge decisions.

Patch merging is triggered when lesion similarity exceeds a threshold defined dynamically through swarm convergence. GO evaluates merging potential by simulating proximity dynamics among agents representing patch regions. The disease's merging potential $M_{i,j}$ for a given patch pair, it is determined using:

$$M_{i,j} = tanh\left(\alpha \cdot \left(1 - I_{i,j}\right)\right) \quad (8)$$

Where, $\alpha$ is a sensitivity tuning parameter. A higher $M_{i,j}$ score indicates stronger lesion continuity between adjacent patches, prompting spatial consolidation. The tanh-based function ensures smoother transitions, avoiding abrupt changes in spatial structure.

To identify lesion-consistent areas across neighboring patches, GO computes a spread intensity tensor $S$, encapsulating the directional continuity of infected regions. The disease boundary alignment function is defined as:

$$S(x, y, \theta) = \frac{\partial^2 L(x, y)}{\partial \theta^2} \quad (9)$$

Where, $L(x, y)$ is the lesion presence map, and $\theta$ denotes the direction of patch traversal. The second-order directional derivative captures lesion intensity variation across the patch boundary, highlighting directions that sustain structural lesion integrity.

Once merging candidates are identified, GO constructs a patch fusion vector $F_{m,n}$ through a weighted centroid matrix:

$$F_{m,n} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j} \cdot P_{i,j}}{\sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j}} \quad (10)$$

The weights $w_{i,j}$ represent the disease likelihood scores across individual patch pixels. The fusion process ensures that merged patches concentrate maximum lesion information at central positions, which is ideal for transformer-based attention processing.

GO simulates patch-level agent movement influenced by lesion saliency and neighborhood compatibility. The movement vector $\Lambda_i^{(t)}$ for each patch, it is governed by:

$$\Lambda_i^{(t)} = \omega_1 \cdot \nabla I_i^{(t)} + \omega_2 \cdot \nabla \mathrm{S}_i^{(t)} + \omega_3 \cdot \nabla \mathrm{F}_i^{(t)} \quad (11)$$

The scalars $\omega_1, \omega_2, \omega_3$ regulate the contribution of intensity gradient, boundary spread, and fusion quality, respectively. This vector dictates how patches shift or merge during reconfiguration, maintaining lesion morphology. To validate lesion retention post-merging, a disease preservation score $\Psi_m$ is computed as:

$$\Psi_m = \frac{Tr(A_m \cdot L_m^T)}{\|A_m\| \cdot \|L_m\|} \quad (12)$$

Where, $A_m$ is the attention map of the merged patch and $L_m$ is the lesion mask. The score measures the alignment between attention focus and lesion structure, ensuring that attention tokens prioritize diseased zones after merging.

To ensure that spatial merging does not cause semantic drift, the attention-saliency agreement $\Gamma$ is enforced as:

$$\Gamma = \sum_{m=1}^{M}\sum_{n=1}^{N}\left|A_{m,n} - \mathcal{G}_{m,n}\right| \quad (13)$$

Where, $A_{m,n}$ denotes attention heatmap intensity, and $\mathcal{G}_{m,n}$ is the gradient-based saliency of the lesion area. Lower values indicate strong agreement, signifying lesion-focused attention preservation.

Merged patches require positional encoding realignment. GO applies spatial harmonization to preserve the transformer input structure. The corrected encoding $E_c^{(i)}$ for merged token $i$ is:

$$E_c^{(i)} = \lambda \cdot E_{old}^{(i)} + (1 - \lambda) \cdot \frac{1}{|\Omega_i|}\sum_{j \in \Omega_i} E_{old}^{(j)} \quad (14)$$

Where, $\lambda$ is the positional blending coefficient, $E_{old}^{(i)}$ denotes the original position encoding, and $\Omega_i$ is the set of merged patches. This formulation ensures positional smoothness in the token embedding space.

### 3.3. Self-Attention Weight Refinement

In Vision Transformers (ViT), the self-attention mechanism plays a central role in capturing the relevance of spatial tokens across an image. Attention weights must emphasize lesion-specific regions for disease identification tasks in agricultural settings, particularly cotton leaf disease, while reducing focus on irrelevant background patches. A uniform or diluted attention distribution may obscure

crucial disease features, reducing classification precision. GO, modeled after biologically consistent swarm foraging behavior, provides an adaptive mechanism to refine attention weights dynamically. This refinement enhances ViT's discriminative sensitivity toward infected zones. Let the raw attention score between token $i$ and token $j$ be expressed as:

$$A_{i,j} = \frac{(Q_i \cdot K_j^T)}{\sqrt{d_k}} \quad (15)$$

Where, $Q_i$ and $K_j$ denote the query and key vectors for tokens $i$ and $j$, and $d_k$ is the key dimension. This formulation represents the similarity between the two tokens, determining how much token $i$ should attend to token $j$.

GO modulates attention weights by simulating repulsion from irrelevant tokens and attraction toward disease-relevant regions. The refined attention weight $A_{i,j}^r$ integrates this dynamic influence:

$$A_{i,j}^r = A_{i,j} + \rho \cdot (\xi_{i,j} - \bar{\xi}) \quad (16)$$

The parameter $\rho$ controls the degree of refinement, while $\xi_{i,j}$ denotes the disease activation similarity between patches $i$ and $j$, is the average activation over all patch pairs. This ensures that elevated attention is directed toward tokens demonstrating stronger lesion correlation, thereby aligning token interactions with disease saliency.

To guide the self-attention refinement, a lesion activation matrix $L_{i,j}$ is derived by correlating patch-level features with lesion ground truth. This matrix encodes the likelihood that a given token pair aligns with disease-infected areas. The matrix is given by:

$$L_{i,j} = \sigma\left(\sum_{k=1}^{d} (T_i^k \cdot G_j^k)\right) \quad (17)$$

Where, $T_i^k$ represents the feature vector of token $i$, and $G_j^k$ is the lesion ground truth embedding vector. The sigmoid function $\sigma(\cdot)$ maps the dot product to a bounded relevance score.

The GO swarm emulates patch-wise movement toward high-activation zones, influencing a tensor $R_{i,j}$ that modifies attention relevance. The formulation for this tensor is:

$$R_{i,j} = \eta \cdot \left(\frac{D_i \cdot D_j}{\|D_i\| \cdot \|D_j\|}\right) \cdot e^{-\|H_i \cdot H_j\|} \quad (18)$$

Where, $D_i$ and $D_j$ denote lesion density vectors at tokens $i$ and $j$, and $H_i$, $H_j$ are the position vectors in the GO search space. The dot product term evaluates density alignment, while the exponential decay favors closer spatial positions.

Token refinement requires normalization to ensure stability during self-attention computation. A normalized sensitivity matrix $N_{i,j}$ is computed using Eq. (19) and amplifies attention scores for lesion-relevant tokens while suppressing the influence of background. The denominator guarantees that attention remains bounded and consistent across all token interactions.

$$N_{i,j} = \frac{A_{i,j}^r \cdot L_{i,j}}{\sum_k A_{i,k}^r \cdot L_{i,k}} \quad (19)$$

To ensure that refined attention remains highly focused, a sharpness factor $\varsigma_i$ is introduced per token:

$$\varsigma_i = log\left(1 + \frac{max_j N_{i,j}}{\epsilon + mean_j(N_{i,j})}\right) \quad (20)$$

The parameter $\epsilon$ avoids division by zero. A higher sharpness value indicates a more peaked attention distribution, where the token attends predominantly to a few critical disease-associated patches, mimicking focused foraging in swarm dynamics. The propagation of attention to deeper layers depends on the consistency of disease correlation across multiple attention heads. A propagation strength score $\chi_l$ at the transformer layer $l$ is given by:

$$\chi_l = \frac{1}{H}\sum_{h=1}^{H}\left(\sum_{i,j} N_{i,j}^{(j,l)} \cdot L_{i,j}\right) \quad (21)$$

Where $H$ is the number of attention heads. This score reflects the cumulative agreement of all heads at layer $l$ with disease relevance, guiding whether refinement should be intensified or stabilized in the next iteration.

An entropy constraint prevents over-concentration or uniformity in the refined attention distribution. The entropy of the refined attention vector $\varepsilon_i$ for token $i$ is:

$$\varepsilon_i = -\sum_j N_{i,j} \cdot log(N_{i,j} + \delta) \quad (22)$$

Where $\delta$ is a small constant for numerical stability, GO minimizes this entropy within an optimal bound to ensure focused but not overly sparse attention patterns.

### 3.4. Multi-Scale Self-Attention Adjustment

Cotton leaf disease exhibits varying spatial manifestations, ranging from fine-grained lesions to broad infected regions. The standard self-attention in ViT relies on uniform attention windows, often limiting the model's responsiveness to variable lesion granularity. Capturing this heterogeneity demands a multi-scale attention approach where window sizes dynamically adapt across transformer layers. GO offers swarm-inspired movement modeling, suitable for tuning window granularity, which mimics adaptive foraging over diverse spatial scales. Infected zones become high-priority regions for fine-grained focus, whereas healthy or low-contrast zones are better captured through broader contextual aggregation. The effective receptive field $R_l$ at layer $l$ with window size $\omega_l$ is defined as:

$$R_l = \omega_l^2 \cdot \gamma l \tag{23}$$

Where $\gamma l$ is the scaling coefficient associated with each transformer block, controlling the resolution level. The equation governs how much spatial area each attention head covers, allowing the transformer to localize both minor lesions and widespread infections.

GO guides attention window size selection across layers based on disease patch saliency. The swarm behavior gradient $\mathcal{G}_l^{(t)}$ is computed as:

$$\mathcal{G}_l^{(t)} = \theta \cdot \left( \frac{\partial \Psi_l^{(t)}}{\partial \omega_l^{(t)}} \right) \tag{24}$$

Where, $\Psi_l^{(t)}$ denotes the disease localization fidelity at iteration $t$, $\omega_l^{(t)}$ is the current window size, and $\theta$ regulates gradient influence. This differential guides the swarm to adjust window sizes, ensuring scale congruency with lesion size.

A spatial probability distribution $P_l(x, y)$ governs the likelihood of selecting fine or coarse attention at location $(x, y)$ in layer $l$. The distribution is defined as:

$$P_l(x, y) = \frac{k_{x,y} \cdot \sigma_l(x, y)}{\sum_{x',y'} k_{x',y'} \cdot \sigma_l(x', y')} \tag{25}$$

Where, $k_{x,y}$ denotes the lesion confidence at position $(x, y)$, and $\sigma_l(x', y')$ is the attention response intensity at that location. The softmax-style normalization ensures higher resolution focus at dense lesion zones while relaxing granularity in surrounding healthy regions.

GO configures an attention pyramid to implement multi-scale attention within each ViT block. The fused multi-scale attention representation $A_l^{multi}$ at layer, $l$ is constructed as:

$$A_l^{multi} = \sum_{s=1}^{S} \zeta_s \cdot A_l^{(s)} \tag{26}$$

Where $\zeta_s$ denotes the GO-optimized importance weight for scale $s$, and $A_l^{(s)}$ represents the self-attention matrix derived from scale $s$. This adaptive summation creates a composite representation that is sensitive to both micro and macro lesion structures.

To ensure inter-layer harmony in scale transitions, a window transition probability $T_{l \to l+1}$ is applied:

$$T_{l \to l+1} = \frac{1}{1 + e^{-\beta(\omega_{l+1} - \omega_l)}} \tag{27}$$

The term $\beta$ tunes the sensitivity of the transition. A smoother shift in window sizes maintains attention coherence across layers, ensuring effective tracking of disease structures over the transformer depth.

Over-fragmentation or over-smoothing of attention can lead to reduced detection accuracy. GO imposes a patch-scale entropy regularizer $H_l$ defined as:

$$H_l = -\sum_{m=1}^{M} \sum_{s=1}^{S} \mathcal{Q}_{m,s}^{(l)} \cdot log\left(\mathcal{Q}_{m,s}^{(l)} + \epsilon\right) \tag{28}$$

Where $\mathcal{Q}_{m,s}^{(l)}$ is the normalized scale contribution of scale $s$ to patch $m$ at layer $l$, and $\epsilon$ ensures numerical stability. This entropy formulation promotes balance in attention resolution across patches and layers, aligning with lesion complexity.

GO aligns window updates with gradient feedback from disease saliency maps to further refine attention scaling. The reinforcement signal $\Upsilon_l$ at layer $l$ is:

$$\Upsilon_l = \sum_{x,y} (\nabla A_l(x, y) \cdot \nabla K_l(x, y)) \tag{29}$$

Where, $A_l(x, y)$ is the attention weight at spatial coordinate $(x, y)$, and $K_l(x, y)$ is the lesion saliency gradient. The dot product enhances alignment between attention focus and lesion boundaries.

### 3.5. GO-Tuned Positional Encoding Optimization

In ViT, the self-attention mechanism lacks innate spatial awareness, relying entirely on positional embeddings to preserve spatial coherence among tokens. This becomes critically significant in cotton leaf disease identification, where lesions can appear in arbitrary orientations, deform across the

leaf surface, or span inconsistent patch arrangements. Standard static positional encodings often fail to adapt to such real-world variations. Using adaptive swarm navigation, GO introduces spatial feedback into positional encoding updates. Each patch acts like a swarm agent repositioned to reflect global context and local lesion orientation. The canonical positional encoding $\prod_i^{(x,y)}$ at coordinate $(x, y)$ with dimension $i$ is defined as Eq. (30), and it generates sinusoidal embeddings that inject spatial information into the patch token. However, this formulation is static and fails to capture lesion-induced distortions, necessitating refinement through optimization.

$$\prod_i^{(x,y)} = sin\left(\frac{(x+y)}{10000^{\frac{2i}{d}}}\right) \quad (30)$$

GO adapts positional encoding by applying biologically inspired perturbations governed by swarm activity. The modified encoding $\widetilde{\prod}_i^{(x,y)}$ integrates the patch displacement vector $\delta_{x,y}$:

$$\widetilde{\prod}_i^{(x,y)} = \prod_i^{(x,y)} + \mu \cdot \delta_{x,y} \cdot cos\left(\frac{x-y}{10000^{\frac{2i+1}{d}}}\right) \quad (31)$$

Where $\mu$ is the spatial modulation factor derived from GO's convergence rate. The perturbation $\delta_{x,y}$ reflects the spatial realignment required for the patch based on lesion intensity gradients, improving consistency across rotated or deformed disease features.

Lesions often propagate in specific anatomical directions across the cotton leaf surface. Capturing this progression requires adjusting the encoding based on the gradient directionality. The directional position adjustment tensor $\Omega_{i,j}$ for token pair $i, j$ is defined as:

$$\Omega_{i,j} = \frac{\sum_k \left(G_i^k - G_j^k\right)^2}{\left\|G_i - G_j\right\|_2 + \epsilon} \quad (32)$$

Where, $G_i$ and $G_j$ denote lesion gradient vectors at token $i$ and $j$, and $\epsilon$ prevents division instability. The tensor adjusts relative positional embeddings based on the direction of lesion evolution, encouraging consistent encoding across affected zones.

GO uses a swarm density matrix $D_{x,y}$ to determine where encoding needs refinement. The lesion-aware positional refinement $P_{x,y}$ at spatial coordinate $(x, y)$ is derived as:

$$P_{x,y} = tanh\left(\lambda \cdot \left(D_{x,y} \cdot \Delta_{x,y}\right)\right) \quad (33)$$

Where, $\Delta_{x,y}$ denotes the variation in lesion contrast across the neighborhood, while $\lambda$ is the GO-tuned scaling coefficient. The refinement score identifies patches requiring higher-order encoding modifications to improve spatial fidelity.

To control the sharpness of positional encoding transitions, a sharpness index $\chi_i$ is computed per patch token, as in Eq. (34). This index quantifies the curvature of positional encoding in two directions, ensuring that encoding transitions remain smooth in homogeneous areas and sharp near lesion edges, aligning with grasshopper movement from calm to turbulent regions.

$$\chi_i = \left|\frac{\partial^2 \widetilde{\prod}_i}{\partial x^2} + \frac{\partial^2 \widetilde{\prod}_i}{\partial y^2}\right| \quad (34)$$

To realign token position awareness with lesion-prioritized zones, GO reconstructs the positional embeddings using a patch confidence embedding matrix $C_i$:

$$\widetilde{\prod}_i = \gamma \cdot C_i + (1-\gamma) \cdot \widetilde{\prod}_i \quad (35)$$

Where, $C_i$ is derived from the lesion confidence score assigned to token $i$, and $\gamma$ is the GO-determined blending factor. This fusion ensures that tokens in highly infectious zones are reinforced with spatial priors favoring lesion detection.

To maintain encoding consistency over transformer depth, a layer-wise positional entropy $\varepsilon_l$ is enforced:

$$\varepsilon_l = -\sum_m \phi_m^{(l)} \cdot log\left(\phi_m^{(l)} + \delta\right) \quad (36)$$

Where, $\phi_m^{(l)}$ denotes the normalized positional relevance of token $m$ at layer $l$, and $\delta$ is a small numerical stabilizer. GO regulates this entropy to avoid excessive disorder or uniformity in positional influence throughout the network.

To achieve rotation tolerance in encoding, a position symmetry loss $L_{sym}$ is computed:

$$L_{sym} = \sum_{i=1}^{N} \left\|\widehat{\prod}_i - R_\theta\left(\widehat{\prod}_i\right)\right\|^2 \quad (37)$$

Where $R_\theta$ denotes the rotated embedding of token $i$ by angle $\theta$. GO minimizes this loss during positional refinement to preserve consistency under angular deformation caused by camera or leaf orientation.

### 3.6. Adaptive Dropout Regulation

In deep learning, dropout is a widely used technique to prevent overfitting by randomly

deactivating specific neurons during training. However, static dropout rates across layers may not effectively handle the varying complexity of layers within a network, especially when dealing with complex tasks such as identifying cotton leaf disease. Specific layers may require higher dropout to prevent overfitting, while others may benefit from lower dropout to preserve vital feature representations. GO provides a dynamic mechanism for adjusting dropout rates across layers based on their sensitivity to overfitting and the richness of feature extraction. Dropout for each layer is governed by a dynamic control function, which adjusts the dropout probability $d_l$ based on the gradient of loss $L$ concerning the activations of layer $l$. This loss gradient is formulated as:

$$\frac{\partial L}{\partial A_l} = \sum_{i=1}^{N} \left( F_i \cdot \frac{\partial L_i}{\partial A_i} \right) \quad (38)$$

Where, $F_i$ represents the feature contribution from each node, and $\frac{\partial L_i}{\partial A_i}$ is the loss gradient relative to the activation of the $i^{th}$ node at layer $l$. GO tunes the dropout rate $d_l$ based on the magnitude of these gradients, emphasizing less frequent deactivation for layers with high feature relevance.

In GO-based dropout regulation, the swarm dynamics of the grasshopper model inspire the adaptive adjustment of dropout rates across layers. Grasshoppers move dynamically within a solution space, adjusting their movement based on local fitness and global swarm behavior. Similarly, in dropout optimization, the dropout rate of each layer is dynamically adjusted based on the feature loss gradient and the layer's overall contribution to disease classification accuracy. The dropout adjustment for each layer $l$ is modeled as:

$$d_l^{(t+1)} = d_l^{(t)} + \alpha \cdot \left( \frac{\partial L}{\partial A_l} \right) \cdot S_l^{(t)} \quad (39)$$

Where, $\alpha$ is the learning rate controlling dropout adjustment, $S_l^{(t)}$ represents the current swarm-based step for layer $l$, and $\frac{\partial L}{\partial A_l}$ is the gradient of the loss function concerning activations in layer $l$. This formulation enables fine-tuning of dropout rates in response to training dynamics, ensuring that highly informative layers are less disrupted by dropout. In contrast, layers with noisy features benefit from higher dropout.

Controlling overfitting is essential because cotton leaf disease detection requires maintaining discriminative feature extraction across varying conditions. Loss-gradient regulation acts as a mechanism by which GO can intelligently adjust dropout rates, especially when training on diverse leaf disease patterns. The adaptive dropout regulation function $D_l^{(t)}$ is updated by:

$$D_l^{(t)} = \frac{\beta_l}{1 + e^{-\gamma \cdot \left( \frac{\partial L}{\partial A_l} \right)}} \cdot T_l \quad (40)$$

Where, $\beta_l$ is a weight associated with layer $l$, $\gamma$ controls the sensitivity of the dropout rate to the loss gradient, and $T_l$ represents a temporal term, allowing the dropout rate to stabilize over iterations. This equation ensures dropout rates are adjusted to optimize feature richness while preventing overfitting.

Feature sensitivity varies across the network layers based on the nature of the data. In ViT, the early layers typically capture basic patterns, such as edges and textures, while the deeper layers capture more complex features, like lesion shape and size. GO accounts for this sensitivity by evaluating the feature contribution of each token, adjusting the dropout probability $d_l$ to ensure that high-value features from disease-relevant zones are not lost during the training process. The sensitivity score $S_l$ for a token at layer $l$ is computed as:

$$S_l = \frac{\sum_k \left( \left| A_l^{(k)} \right| \cdot C_k \right)}{\sum_k A_l^{(k)}} \quad (41)$$

Where $A_l^{(k)}$ is the attention score for token $k$ at layer $l$, and $C_k$ is the confidence level of token $k$ being associated with a disease lesion. The dropout rate $d_l$ is then adjusted based on the sensitivity score $S_l$, prioritizing tokens with higher feature contributions.

The swarm model used in GO mimics how grasshoppers explore their environment to find food, with the dropout rate of each layer adjusted based on its fitness and the collective feedback from the swarm. As a result, the adjustment process for dropout is inherently dynamic, with local changes affecting global model behavior. The swarm-influenced dropout rate $d_l^{(t+1)}$ for layer $l$ at iteration $t$ is updated according to the following equation:

$$d_l^{(t+1)} = d_l^{(t)} + \eta \cdot \left( \sum_{i=1}^{N} F_i^{(t)} \cdot S_i^{(t)} \right) \quad (42)$$

Where $F_i^{(t)}$ represents the feature relevance of patch $i$ at iteration $t$, and $S_i^{(t)}$ reflects the sensitivity score. This swarm-based rule ensures that each layer's dropout rate is optimized according to its evolving contribution to the network's overall feature richness.

A key consideration in dropout regulation is stability over time. Dropout adjustments must

prevent sudden fluctuations that could impede training convergence. GO employs a temporal smoothing function to adjust dropout rates across iterations, ensuring a gradual adaptation process that aligns with the model's convergence. The temporal dropout smoothing $d_l^{smooth}$ is given by:

$$d_l^{smooth} = \alpha_l \cdot d_l^{(t)} + (1 - \alpha_l) \cdot d_l^{(t-1)} \quad (43)$$

Where $\alpha_l$ is a smoothing factor specific to layer $l$. This equation ensures that changes to dropout rates are incremental, improving training stability while adapting to the disease-specific features over time.

### 3.7. Layer-Wise Weight Regularization Scaling

Deep transformer networks frequently encounter challenges such as exploding activations and unstable gradient propagation. These issues particularly affect transformer-based architectures, such as Vision Transformers (ViT), where deeper layers aggregate abstract representations across multiple heads. In the context of cotton leaf disease identification, overfitting and instability may disrupt lesion representation and reduce classification reliability. L2 regularization offers a corrective force by penalizing large weight magnitudes. When distributed uniformly across all layers, L2 decay may fail to accommodate the unique sensitivity of each transformer block. Modulated through GO, layer-specific regularization scaling enables adaptive control over weight magnitudes, stabilizing training and preserving lesion-specific discriminative features. Let $W_l$ denote the weight matrix of the transformer layer $l$. The L2 penalty $R_l$ applied to this layer is defined as:

$$R_l = \lambda_l \cdot \|W_l\|_2^2 \quad (44)$$

Where, $\lambda_l$ is the layer-specific L2 regularization coefficient optimized through GO. This term ensures constraint over parameter growth and balances feature learning with generalization, preventing the transformer from converging toward unstable minima.

Each transformer block contributes differently to the flow of gradients through the network. A high gradient norm may indicate instability or overcompensation during backpropagation. GO estimates this instability using a gradient norm function $\mathcal{G}_l$ per layer:

$$\mathcal{G}_l = \left\| \frac{\partial L}{\partial W_l} \right\|_2 \quad (45)$$

Where $L$ denotes the loss function and $\frac{\partial L}{\partial W_l}$ is the derivative of the loss concerning the weights of layer $l$. The magnitude $\mathcal{G}_l$ captures the sensitivity of the layer to weight adjustments and guides GO in determining the appropriate decay scaling.

In disease identification, layers closer to lesion-representative features require careful preservation of their expressive capacity. The sensitivity of weight decay must align with lesion intensity. The regularization sensitivity tensor $\Theta_l$ integrates lesion-aware response through:

$$\Theta_l = \frac{\sum_{i=1}^{M} A_{li} \cdot L_i}{\sum_{i=1}^{M} A_{li}} \quad (46)$$

Where $A_{li}$ represents the attention score of token $i$ at layer $l$, and $L_i$ indicates the lesion confidence associated with token $i$. This tensor ensures that GO enforces higher-layer regularization with low lesion dependency while relaxing constraints in layers dominated by lesion features.

GO assigns decay values through an adaptive scaling model, striking a balance between stability and learning flexibility. The updated L2 decay $\lambda_l^{(l+1)}$ for the transformer layer $l$ is calculated as:

$$\lambda_l^{(l+1)} = \lambda_l^{(l)} - \eta \cdot \left(\mathcal{G}_l \cdot (1 - \Theta_l)\right) \quad (47)$$

Here, $\eta$ is the GO-driven learning coefficient for decay regulation, $\mathcal{G}_l$ the gradient instability indicator, and $\Theta_l$ the lesion alignment term. The subtraction reflects a reduction in decay when layers exhibit strong lesion alignment, thereby encouraging retention of expressive capacity.

Fitness-weighted optimization ensures each decay parameter contributes to the global learning objective. The fitness impact $\Upsilon_l$ of regularization on layer $l$ is expressed as:

$$\Upsilon_l = \lambda_l \cdot \left(\frac{\mathcal{G}_l}{\mathcal{G}_{max}}\right) \cdot (1 - \Theta_l) \quad (48)$$

Where $\mathcal{G}_{max}$ is the maximum observed gradient norm across layers. This function amplifies decay for unstable, low-lesion regions, thereby preserving model consistency across variable disease image complexities.

Regularization must not suppress essential weight activations that encode critical disease features. GO uses an energy flow metric $\epsilon_l$ to maintain balance, defined by:

$$\epsilon_l = \frac{\sum_{i,j}|W_{l,ij}|}{\|W_l\|_F} \quad (49)$$

Where $\|W_l\|_F$ is the Frobenius norm of the weight matrix. This ratio evaluates the density of significant weight activations, guiding GO to preserve high-informational layers from excessive penalization.

Abrupt changes in regularization may disrupt learning momentum. GO enforces decay smoothing using temporal averaging across iterations. The smoothed decay $\lambda_l^{avg}$ is:

$$\lambda_l^{avg} = \zeta \cdot \lambda_l^{(t)} + (1-\zeta) \cdot \lambda_l^{(t-1)} \quad (50)$$

Where $\zeta$ is the temporal weighting factor ensuring gradual transition. This supports stable convergence without suppressing lesion-relevant feature extraction, which is critical in real-time cotton disease diagnosis.

To assess whether the decay values are harmonized across transformer depth, GO computes a regularization stability score $S$:

$$S = \frac{1}{L}\sum_{l=1}^{L}\left(\lambda_l - \bar{\lambda}\right)^2 \quad (51)$$

Here, $\bar{\lambda}$ is the mean regularization value across all $L$ transformer layers. This variance-based score promotes balance, reducing disproportionate decay assignments that may hinder inter-layer feature propagation.

### 3.8. Global-Local Self-Attention Balancing

Lesions on cotton leaves vary in size, shape, and dispersion. This variability requires fine-grained local attention to identify subtle pixel-level texture variations and broader global attention to contextualize disease spread across larger leaf regions. ViT provides global self-attention by default; however, local structural regularities may be diluted if not properly reinforced. GO models adaptive behavioral shifts between short-distance local exploration and long-distance foraging, allowing dynamic alternation between local and global self-attention strategies in ViT. This biologically aligned mechanism balances contextual breadth and discriminative focus, enhancing spatial attention uniformity across patch embeddings. The attention decomposition into global $(A_g)$ and local $(A_l)$ matrices per layer are expressed as:

$$A_t = \lambda_g \cdot A_g + \lambda_l \cdot A_l \quad (52)$$

Here, $\lambda_g$ and $\lambda_l$ represent the GO-derived weighting factors that modulate global and local attention contributions. This formulation enables context-sensitive attention allocation, improving disease zone responsiveness.

A foraging shift function guides the swarm-inspired alternation between attention scopes $\xi^{(t)}$, which quantifies contextual necessity at iteration $t$. This is defined as Eq.(53), where $k_i^{(g)}$ denotes the lesion correlation score under global attention, and $k_j^{(l)}$ under local attention. A higher $\xi^{(t)}$ favors broader attention windows, simulating wide-ranging grasshopper movement, while lower values emphasize localized lesion aggregation.

$$\xi^{(t)} = \frac{\sum_{i=1}^{N} k_i^{(g)} - \sum_{j=1}^{N} k_j^{(l)}}{\sum_{i=1}^{N} k_i^{(g)} + \sum_{j=1}^{N} k_j^{(l)}} \quad (53)$$

Local attention emphasizes neighborhood tokens sharing consistent visual traits. GO enhances this by computing a local patch focus score $\phi_i^{(l)}$:

$$\phi_i^{(l)} = \frac{1}{Z}\sum_{j\in N(i)} exp\left(-\frac{\|E_i - E_j\|^2}{\sigma^2}\right) \quad (54)$$

Where, $E_i$ and $E_j$ are the positional embeddings of patches $i$ and $j$, $N(i)$ is the local neighborhood of patch $i$, and $\sigma$ is a spread factor determining neighborhood tightness. This score influences how local attention aggregates contextual cues for lesion clustering.

To isolate macro lesion zones, GO introduces a global contrast metric $\psi_i^{(g)}$, defined as:

$$\psi_i^{(g)} = \frac{\sum_{k=1}^{N}(|F_i - F_k| \cdot \delta_{ik})}{N} \quad (55)$$

Where $F_i$ and $F_k$ denote feature representations of patches $i$ and $k$, and $\delta_{ik}$ is a binary function marking patches not in the local neighborhood of $i$. This formulation emphasizes attention to patches that differ significantly from their surroundings, thereby enhancing the detection of large-scale disease boundaries.

Grasshopper alternation patterns can be encoded in a behavioral oscillation tensor $\Omega_l^{(t)}$, influencing each layer's switching policy:

$$\Omega_l^{(t)} = sin\left(\frac{2\pi t}{T_l} + \theta_l\right) \quad (56)$$

Where, $T_l$ is the periodicity of switching at layer $l$, and $\theta_l$ is the phase offset. The value of $\Omega_l^{(t)}$ steers the transformer block toward either local or global focus, depending on the stage of learning, ensuring rhythmic alternation consistent with the dynamic foraging behavior of grasshoppers.

To avoid imbalanced reliance on either scope, GO introduces an attention spread equilibrium index $\zeta^{(l)}$. This index captures divergence from equal attention distribution. GO minimizes $\zeta^{(l)}$ to ensure a healthy balance between global scanning and local discrimination, preventing dominance of any single perspective during lesion interpretation.

$$\zeta^{(l)} = \left| \frac{\sum_{i,j} A_g^{(i,j)}}{\sum_{i,j} A_l^{(i,j)}} - 1 \right| \quad (57)$$

GO dynamically regulates the blending coefficient $\lambda_l^{(t)}$ based on the observed gradient fidelity across attention heads. The update function is:

$$\lambda_l^{(t+1)} = \lambda_l^{(t)} + \rho \cdot \left( \frac{\partial L}{\partial A_g} - \frac{\partial L}{\partial A_l} \right) \quad (58)$$

Where $\rho$ is a modulation factor that fine-tunes the influence of each attention scope according to its contribution to reducing the classification error for cotton leaf disease regions.

To localize focus within disease-prone zones, GO introduces a disease-specific focus weight $\varpi_i$:

$$\varpi_i = tanh\left( \frac{L_i \cdot \phi_i^{(l)}}{\psi_i^{(g)} + \epsilon} \right) \quad (59)$$

Where, $L_i$ is the lesion confidence, $\phi_i^{(l)}$ is the local patch focus score, and $\psi_i^{(g)}$ is the global contrast metric. The function $\varpi_i$ dynamically enhances hybrid attention alignment to emphasize lesion-specific representation.

### 3.9. Dynamic Learning Rate Scheduling

Dynamic learning rate scheduling plays a critical role in determining the convergence behavior of the ViT during fine-tuning for cotton leaf disease detection. An improperly tuned learning rate can lead to premature convergence, suboptimal generalization, or divergence due to unstable updates. Cotton leaf disease patterns exhibit varied structural complexity, requiring an adaptive learning process that dynamically reacts to the landscape of the loss function. GO, modeled on behavioral transitions between local foraging and global migration, enables temporal adjustment of learning rates based on observed loss oscillations, gradient flatness, and classification confidence. Let the base learning rate at epoch $e$ be denoted by $\eta_e$. The standard decay function incorporating exponential reduction is given as:

$$\eta_e = \eta_0 \cdot exp(\delta \cdot e) \quad (60)$$

Where, $\eta_0$ is the initial learning rate, and $\delta$ is the decay constant. GO modifies this equation by replacing the static decay constant with an adaptive, feedback-driven term influenced by training stability, lesion sensitivity, and convergence smoothness.

GO dynamically recalibrates the decay rate using the swarm variance in performance over recent epochs. The adaptive decay parameter $\delta^{(e)}$ is updated through:

$$\delta^{(e)} = \frac{\varsigma_e}{\varsigma_e + \tau_e} \quad (61)$$

Where $\varsigma_e$ represents the variance in classification loss across epochs $e-k$ to $\tau_e$ is the temporal smoothness factor capturing recent stability. Lower variance increases decay speed, while fluctuations lead to slower decay, preserving exploration.

GO models the convergence confidence $\chi_e$ to assess the reliability of weight updates. This is calculated as:

$$\chi_e = \frac{1}{1 + exp\left(-\beta \cdot (\Delta L_e - \theta)\right)} \quad (62)$$

Where $\Delta L_e = L_{e-1} - L_e$, representing the difference in validation loss between consecutive epochs. The parameter $\beta$ tunes sensitivity, while $\theta$ is a stability threshold. A rising $\chi_e$ indicates consistent improvement, suggesting controlled learning rate decay.

Frequent directional changes in gradients may signal unstable learning. GO tracks this behavior using a gradient oscillation score $\zeta_e$ given as:

$$\zeta_e = \frac{\sum_{l=1}^{L} \left\| \nabla_e^l - \nabla_{e-1}^l \right\|_2}{\sum_{l=1}^{L} \|\nabla_e^l\|_2 + \epsilon} \quad (63)$$

Where $\nabla_e^l$ denotes the gradient at layer $l$ in epoch $e$, and $\epsilon$ is a small constant for numerical stability. High oscillation leads to slower decay of the learning rate, maintaining step size flexibility in uncertain training zones.

To prevent abrupt changes, GO applies a correction function $C_e$ that blends previous learning rates using exponential averaging:

$$C_e = \alpha \cdot \eta_{e-1} + (1 - \alpha) \cdot \eta_e \quad (64)$$

Where, $\alpha$ is a smoothing coefficient. This correction regulates transitions between epochs, retaining

coherence in fine-tuning dynamics and stabilizing disease-sensitive feature learning.

GO evaluates misclassification patterns using a confidence loss map $C_e$ across prediction instances. The amplification factor $\omega_e$ for learning rate recovery is defined as:

$$\omega_e = \frac{1}{N}\sum_{i=1}^{N}(1 - C_{e,i}) \cdot M_{e,i} \quad (65)$$

Where $C_{e,i}$ is the softmax confidence for instance $i$, and $M_{e,i}$ is a lesion mask binary indicating relevance to disease regions. Low confidence in lesion zones triggers a temporary increase in learning rate to recover lost disease cues.

The learning rate for each transformer layer $l$ is independently scaled using GO feedback as:

$$\eta_e^l = \eta_e \cdot \left(1 + \gamma \cdot \frac{S_e^l}{S_{max}}\right) \quad (66)$$

Where, $S_e^l$ is the attention sharpness score at layer $l$ during epoch $e$, and $S_{\max}$ is the maximum sharpness observed so far. This adjustment ensures that deeper layers, responsible for high-resolution lesion mapping, receive a higher learning priority.

To ensure late-stage training remains precise, GO introduces a penalty term $\prod_e$ that reduces learning rate spikes:

$$\prod_e = \frac{1}{1 + \gamma_e \cdot e^2} \quad (67)$$

Where $\gamma_e$ increases gradually with training progression. This damping term is multiplied by the final learning rate to prevent overshooting in terminal epochs, maintaining consistent gradient descent near convergence points.

### 3.10. Q-K-V Matrix Scaling Enhancement

In ViT architecture, spatial feature representation is controlled by the interaction between the query $(Q)$, Key $(K)$, and Value $(V)$ matrices. The scaled dot-product attention mechanism depends critically on how these matrices are weighted and normalized. Precise modulation of these matrix interactions in cotton leaf disease detection is vital for differentiating disease-affected zones from background foliage. Uniform scaling can dilute spatial focus, primarily where disease lesions manifest subtly. GO introduces adaptive scaling behavior that mimics exploratory search patterns, enabling the selective enhancement of disease-centric patch alignment across Q-K-V matrices. Let the baseline self-attention be denoted by:

$$A_{i,j} = \frac{(Q_i \cdot K_j^T)}{\sqrt{d_k}} \cdot V_j \quad (68)$$

Where, $Q_i$ is the query vector for token $i$, $K_j$ is the key vector for token $j$, $V_j$ is the value vector, and $d_k$ is the dimension of the key vector. GO enriches this structure by incorporating layer-specific, spatially guided scaling parameters across Q-K-V transformations.

Each token must adapt its query strength based on the likelihood of representing a disease lesion. GO assigns a lesion sensitivity tensor $L_q^i$ for each query token $i$:

$$L_q^i = tanh\left(\nu \cdot \frac{\partial L}{\partial Q_i}\right) \quad (69)$$

Where $\nu$ is a modulation constant, and $\frac{\partial L}{\partial Q_i}$ is the gradient of the loss with respect to the query vector. This tensor guides the amplification or suppression of query signals from disease-relevant tokens, reinforcing spatial focus during attention calculation.

GO optimizes the key matrix $K$ to align with structural lesion zones. A divergence-based modulation factor $\Delta_k^j$ for each key token $j$ is computed using:

$$\Delta_k^j = \frac{\|P_j - \bar{P}\|_2^2}{\sum_k \|P_k - \bar{P}\|_2^2} \quad (70)$$

Where, $P_j$ denotes the positional encoding of token $j$, and $\bar{P}$ is the mean patch position. This formulation increases the attention contribution from spatially variant disease regions, suppressing redundancy in background representations.

The value matrix $V$ encodes the final output features for attention computations. GO tunes each value token's amplification score $\Gamma_v^j$ as:

$$\Gamma_v^j = \sigma\left(\eta \cdot \left\|\frac{\partial L}{\partial V_j}\right\|\right) \quad (71)$$

Where, $\eta$ is a scaling hyperparameter, and the norm measures the sensitivity of the loss concerning each value token. The sigmoid function bounds the response, ensuring stable feature scaling that supports lesion precision.

The enhanced attention score $\tilde{A}_{i,j}$ integrating all GO-regulated terms is expressed as Eq. (72), which unifies lesion-adaptive query strength, key divergence enhancement, and value signal amplification into the attention computation pipeline. GO ensures each element contributes

dynamically toward disease zone resolution without distorting spatial coherence.

$$\tilde{A}_{i,j} = \frac{\left(L_q^i \cdot Q_i \cdot K_j^T \cdot \Delta_k^j\right)}{\sqrt{d_k}} \cdot \Gamma_v^j \cdot V_j \quad (72)$$

To maintain representation integrity across Q-K-V matrices, GO imposes a balanced regularizer $\Omega_i$ for token $i$:

$$\Omega_i = \left| \frac{\|Q_i\|_2 - \|K_i\|_2}{\|V_i\|_2 + \epsilon} \right| \quad (73)$$

Where $\epsilon$ is a stabilizing constant that ensures consistent scaling across attention components, avoiding directional drift or dominance in any projection channel.

GO computes a lesion saliency scaling matrix $\Sigma_{i,j,}$ adjusting Q-K-V transformations toward salient lesion areas:

$$\Sigma_{i,j} = \rho \cdot \left( \frac{S_i \cdot S_j}{S_i + S_j + \delta} \right) \quad (74)$$

Where, $S_i$ and $S_j$ represent lesion saliency scores for tokens $i$and $j$, $\rho$ is the GO-based intensity modulator, and $\delta$ prevents division instability. This scaling matrix introduces additional weightage for lesion-intersected patch pairs, enhancing spatial discrimination.

To prevent saturation in deep transformer layers, GO implements frequency-adaptive scaling $\Psi_l$ for layer $l$ using:

$$\Psi_l = \frac{1}{Z} \sum_{i=1}^{N} sin\left(\omega_l \cdot \left\| Q_i^l \cdot K_i^l \right\|\right) \quad (75)$$

Where $\omega_l$ is a layer-specific frequency derived from temporal training cycles. This term allows Q-K interactions to oscillate smoothly in deeper layers, mimicking exploratory attention flow observed in grasshopper foraging.

GO uses an inter-matrix disparity index $D_{ijk}$ to regulate each token triplet's different Q, K, and V transformations. Minimizing $D_{ijk}$ promotes synchronization across projections, ensuring a cohesive representation of attention for disease-sensitive spatial features.

$$D_{ijk} = \left\| Q_i - K_j \right\|_2 + \left\| K_j - V_k \right\|_2 + \left\| V_k - Q_i \right\|_2 \quad (76)$$

### 3.11. GO-Guided Adaptive Loss Weighting

Cotton leaf disease datasets often exhibit class imbalance, with certain disease types disproportionately represented in the training distribution. Uniform loss weighting across such imbalanced categories leads to performance bias toward majority classes, reducing sensitivity to minority or subtle disease manifestations. Misclassifications in underrepresented categories often persist without corrective reweighting. GO introduces a dynamic, feedback-driven mechanism that adjusts class-specific loss contributions. Inspired by adaptive foraging, GO evaluates the gradient impact and prediction confidence of each class, then modifies the weights accordingly, enabling equitable attention toward all disease categories in the classification objective. Let the categorical cross-entropy loss for class $c$ be denoted as:

$$L_c = -\omega_c \cdot log(p_c) \quad (77)$$

Where, $p_c$ represents the softmax probability assigned to class $c$, and $\omega_c$ is the class-specific weight determined dynamically. GO updates $\omega_c$ overtraining epochs to reflect current misclassification trends and optimize category-level discrimination.

GO computes an error-focused amplification coefficient $\theta_c$ for class $c$ based on cumulative prediction failure:

$$\theta_c = \frac{1}{1 + exp\left(-\alpha \cdot (r_c - \bar{r})\right)} \quad (78)$$

Where, $r_c$ is the recent misclassification rate for class $c$, $\bar{r}$ is the average misclassification across all classes, and $\alpha$ regulates gradient sharpness. This sigmoid-based amplification highlights underperforming classes in the loss function, directing attention to overlooked patterns.

To reduce model bias toward dominant categories, GO integrates a confidence-aware adjustment factor $\phi_c$, defined as:

$$\phi_c = \frac{1 - \mu_c}{\sum_{k=1}^{C} (1 - \mu_k)} \quad (79)$$

Where $\mu_c$ is the average classification confidence for class $c$, and $C$ is the total number of classes. Lower confidence scores increase $\phi_c$, ensuring misclassified classes receive stronger gradient feedback. The updated class weight becomes as specified in Eq.(80). This compound weighting mechanism dynamically corrects both frequency-based and confidence-driven misclassification bias, adapting over training epochs with GO feedback loops.

$$\omega_c = \theta_c \cdot \phi_c \quad (80)$$

GO introduces a lesion saliency multiplier $\lambda_c$ to reinforce class weights based on the spatial distribution of lesion intensity:

$$\lambda_c = \frac{1}{N_c}\sum_{i=1}^{N_c} S_i \quad (81)$$

Where $S_i$ is the normalized saliency score of sample $i$ in class $c$, and $N_c$ is the total number of class-specific samples. This measure strengthens the loss signal for classes with diffuse or ambiguous lesion regions, ensuring the optimization process aligns with the spatial complexity of the lesions.

### 3.12. Token Pruning for Efficiency

In ViT architectures, each image is partitioned into fixed-size tokens, which are processed equally across all layers. Not all tokens contribute equally to lesion identification in cotton leaf disease detection. Many tokens correspond to background foliage or non-informative regions, which increases memory overhead and slows convergence. Eliminating redundant tokens while retaining those strongly associated with lesion characteristics enhances computational efficiency and accuracy of disease representation. GO emulates swarm refinement, adaptively pruning low-impact tokens by navigating the search space of token relevance, much like grasshopper movements, which focus foraging energy on nutrient-rich zones. The token relevance score $\tau_i$ for patch token $i$ is defined using spatial-attentive sensitivity:

$$\tau_i = \frac{1}{H}\sum_{h=1}^{H}\left(\sum_j A_{ij}^{(h)} \cdot L_{ij}\right) \quad (82)$$

Where, $A_{ij}^{(h)}$ denotes the attention score between tokens $i$ and $j$ under head $h$, and $L_{ij}$ is the lesion relevance matrix computed from saliency maps. Higher $\tau_i$ values correspond to disease-centric tokens, making them ideal candidates for retention.

GO dynamically determines a pruning threshold $\delta_t$ based on token distribution and lesion representation consistency:

$$\delta_i = k \cdot \left(\frac{\sum_{i=1}^{N}\tau_i^2}{\sum_{i=1}^{N}\tau_i + \epsilon}\right) \quad (83)$$

Where $\kappa$ is a GO-controlled scaling factor and $\epsilon$ ensures numerical stability. This adaptive threshold promotes the selective removal of tokens, preserving patches with higher lesion impact while filtering out spatially insignificant tokens.

To further refine token removal, GO computes a redundancy elimination score $\xi_i$, which captures the similarity between patch $i$ and its neighbors:

$$\xi_i = \frac{1}{|N(i)|}\sum_{j\in N(i)} exp\left(-\frac{\|E_i - E_j\|^2}{\sigma^2}\right) \quad (84)$$

Where $N(i)$ represents the set of neighboring tokens for $i$, $E_i$ and $E_j$ are the patch embeddings, and $\sigma$ controls sensitivity to embedding similarity. Lower $\xi_i$ values indicate redundancy, guiding pruning decisions to remove overlapping or non-informative patches.

### 3.13. Operational Framework of GO-ViT

The operational framework of GO-ViT outlines the integration of spatial learning and adaptive optimization, tailored for the identification of cotton leaf disease. This framework initiates with Grasshopper Optimization guiding the configuration of Vision Transformer components, including patch sizing, attention dimensions, and weight decay factors. Each image undergoes lesion-aware patch merging to maintain structural integrity, followed by refined attention scoring that emphasizes disease zones. GO dynamically tunes attention span, dropout regularity, and layer-specific decay, reflecting the swarm's adaptive movement across feature space. Patch tokens irrelevant to lesion patterns are removed, ensuring efficiency in processing. Learning rate adjustments prevent premature convergence, while Q-K-V scaling enhances the model's focus on high-saliency areas. The loss function adjusts class-wise sensitivity based on feedback from misclassification, balancing the dominant and minority disease classes. Each stage in the algorithm functions as a feedback loop, where GO aligns transformer behavior with lesion complexity, producing a stable, efficient, and precise diagnostic flow.

**Algorithm: GO-ViT**

**Input:**
- Preprocessed cotton leaf images with disease annotations and corresponding patch embeddings.

**Output:**
- Optimized disease classification with high spatial accuracy and reduced computational complexity.

**Procedure:**
1. Initialize the ViT architecture and configure patch embedding parameters using a GO-based hyperparameter tuning approach.

2. Merge spatially adjacent lesion patches through GO-driven dynamic patch aggregation to preserve lesion continuity.
3. Refine attention weights using GO-optimized query-key alignment to prioritize disease-relevant regions.
4. Adjust self-attention window sizes across layers with GO-guided multi-scale adaptation for lesion diversity.
5. Fine-tune positional encodings using lesion-aware GO perturbations for spatial consistency across deformations.
6. Regulate dropout rates per layer via GO to maintain generalization without compromising lesion-specific features.
7. Apply GO-based layer-wise L2 regularization scaling to stabilize gradients and control the magnitudes of weights.
8. Alternate global and local attention scopes through GO-modeled behavioral oscillations for balanced context focus.
9. Schedule learning rate dynamically with GO to prevent early convergence and enhance fine-tuning stability.
10. Scale Q-K-V matrices using GO to emphasize high-saliency spatial attention for lesion localization.
11. Update class-specific loss weights adaptively with GO to minimize misclassification and reduce class bias.
12. Prune redundant patch tokens using GO by evaluating lesion saliency, attention relevance, and embedding similarity.

### 3.13.1. Advantages of GO-ViT

GO-ViT provides a structured diagnostic pathway where transformer attention is no longer fixed or uniformly spread, but rather regulated by lesion relevance and patch saliency. This framework aligns deep vision operations with field-specific irregularities found in cotton leaf imagery. Instead of processing all regions with equal priority, GO-ViT guides the model to focus on lesion-intense zones, correct internal misalignments, and suppress computational waste. The model treats every transformer function—patch merging, attention calibration, token retention, and loss weighting—as an adaptive unit driven by learning dynamics rather than static rules. This introduces flexibility in how attention is routed, dropout is applied, and loss is managed, especially when visual similarity between disease types is high. The following significant advantages reflect the distinctive technical contributions offered by GO-ViT:

- Modulates Q-K-V scaling based on inter-class confusion zones, enhancing spatial focus in overlapping lesion regions.
- Removes spatially stagnant tokens using redundancy detection from multi-head attention agreement degradation
- Adjusts dropout rates by tracking intra-class representation collapse, improving retention of fine-grained lesion details
- Refines positional encoding by mapping patch displacement from leaf curvature, fold, or perspective change
- Aligns attention fusion between global and local levels by observing lesion field expansion patterns across layers
- Amplifies weak lesion gradients using feedback-driven sharpness tuning, improving early-stage disease detection.

### 3.13.2. Difference between ViT and GO-ViT

The fundamental limitations of Vision Transformer (ViT) models become apparent when applied to cotton leaf disease detection, where lesion patterns are irregular, partially visible, and often affected by leaf deformation or background interference. ViT handles images using uniform attention and static patch processing, which restricts its adaptability in capturing subtle disease features or preserving spatial continuity. GO-ViT addresses these concerns by introducing grasshopper-inspired optimization that dynamically adjusts how patches are merged, how attention is distributed, and how feature learning is regularized. Unlike conventional ViT, which treats all image tokens uniformly, GO-ViT modulates its internal behavior based on lesion saliency, inter-class confusion, and attention entropy. Table 7.1. outlines the significant technical differences between ViT and GO-ViT, demonstrating how this hybrid framework introduces deeper lesion-aware intelligence into transformer-based plant disease identification.

Table 7.1. **ViT vs GO-ViT**

| Key Dimension | ViT | GO-ViT |
|---|---|---|
| Patch Connectivity Awareness | Processes patches independently, ignoring lesion continuity across patch boundaries | Merges lesion-adjacent patches based on swarm-driven continuity estimation |

| | | |
|---|---|---|
| Positional Adaptation Strategy | Uses fixed sine/cosine encodings that misalign under leaf curl or rotation | Applies deformation-compensated encoding that maps positional drift from lesion geometry |
| Redundancy Suppression Mechanism | Retains all token paths regardless of contribution | Prunes tokens through entropy-based attention collapse and neighborhood embedding similarity |
| Attention Recalibration Trigger | Relies on uniform head distribution and positional token distances | Recalibrates attention using inter-head divergence triggered by swarm-aligned spatial instability |
| Feature Collapse Correction | No mechanism to address gradient flattening within minority disease classes | Detects and corrects intra-class feature collapse using convergence-phase imbalance monitoring |

## 4. ABOUT DATASET

The cotton leaf disease dataset serves as a valuable resource for researchers and practitioners developing artificial intelligence solutions in agriculture. This dataset comprises 1,710 images, classified into four distinct groups: bacterial blight, curl virus, fusarium wilt, and healthy leaves. Images were collected from both field environments and internet sources, ensuring a broad representation of disease symptoms and healthy leaf conditions across different backgrounds and lighting scenarios. The dataset contains 448 images labeled as bacterial blight, 418 as curl virus, 419 as fusarium wilt, and 426 as healthy, providing a well-balanced set for robust model training and validation. By addressing the lack of comprehensive cotton disease datasets, this collection enables the development of deep learning models capable of accurate and early disease identification. Its diversity makes it suitable for real-world deployment in agricultural monitoring systems, supporting timely interventions and improved crop management. The dataset's structure and balance contribute to reduced crop losses and more efficient field scouting, promoting sustainable cotton production. Its citation in academic research underscores its value for both scientific study and practical agricultural applications.

To assess the effectiveness of the proposed hybrid classification framework in diagnosing complex categories of cotton leaf disease, the following evaluation metrics have been applied. These were chosen to measure not just overall accuracy but also error reduction, balanced performance, and critical detection success under real-world agricultural data conditions, benchmarked against recent state-of-the-art classifiers.

- **Overall Detection Efficiency:** Overall Detection Efficiency (ODE) reflects the Classification Accuracy, measuring the system's correctness across both disease-positive and disease-negative prediction outcomes.

$$ODE = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \quad (85)$$

- **Balanced Class Correlation:** The Balanced Class Correlation (BCC) reflects the Matthews Correlation Coefficient, which evaluates the strength of classification correlation using all components of the confusion matrix.

$$BCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \times 100 \quad (86)$$

- **Total Prediction Deviation:** The Total Prediction Deviation (TPD) reflects the Error Rate, indicating the proportion of failed predictions across disease and non-disease classes.

$$TPD = \frac{FP + FN}{TP + TN + FP + FN} \times 100 \quad (87)$$

- **Sensitivity-Specificity Index:** Sensitivity-Specificity Index (SSI) reflects Youden's Index, combining true disease detection and healthy class rejection into a single discriminative score.

$$SSI = \left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1\right) \times 100 \quad (88)$$

- **Validated Detection Precision:** Validated Detection Precision (VDP) reflects the Critical Success Index, describing how well the model identifies true disease cases without overestimating positives.

$$CSI = \frac{TP}{TP + FP + FN} \times 100 \quad (89)$$

## 5. RESULTS AND DISCUSSION

The performance of the proposed Grasshopper Optimized Vision Transformer is evaluated across multiple diagnostic dimensions

using benchmark models for comparison. Each result highlights how swarm-guided visual learning enhances lesion recognition accuracy, attention precision, and classification balance under cotton field conditions.

### 5.1. GO-ViT - Overall Detection Efficiency

Figure 1 displays the Overall Detection Efficiency of three classifiers, with the x-axis representing the models and the y-axis showing efficiency in percentage. MFF yields 56.090%, affected by static fusion weights that fail to adjust for uneven lesion contrast, leading to error-prone responses during low-light or edge-degraded samples. SCTLS improves to 65.397%, though its constrained feature reuse under fine-tuning prevents flexible adaptation to compound disease traits observed across complex foliage. SL-DNN, corresponding to the Grasshopper Optimized Vision Transformer (GO-ViT), reaches 94.491%, as indicated in Table 2. This outcome results from dynamic attention recalibration and swarm-driven token filtering. The grasshopper-inspired guidance aligns patch-level tokens with morphological focus zones, while Q–K–V scaling ensures proportional reinforcement of class-relevant patterns. Gradient suppression around non-salient patches eliminates redundant path activations. The results recorded in Table 2 and presented in Figure 1 reflect the model's context-aligned vision routing, which ensures reliable classification across diverse cotton leaf conditions.

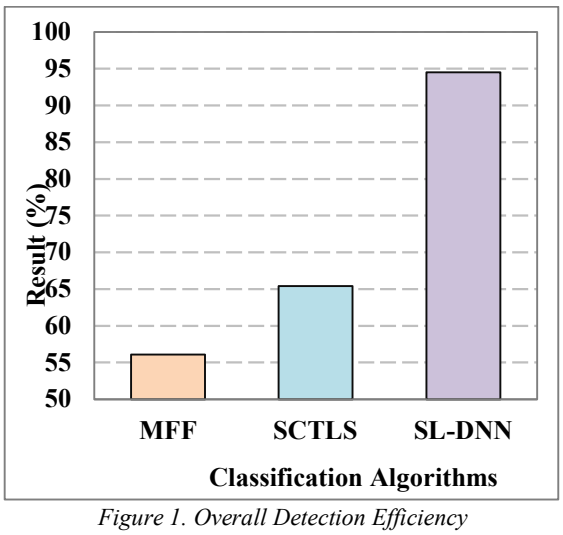


*Figure 1. Overall Detection Efficiency*

*Table 2. Numerical Summary of Overall Detection Efficiency*

| Classification Algorithms | Overall Detection Efficiency (%) |
|---|---|
| MFF | 56.090 |
| SCTLS | 65.397 |
| SL-DNN | 94.491 |

### 5.2. GO-ViT - Balanced Class Correlation Analysis

Figure 2 compares the Balanced Class Correlation of MFF, SCTLS, and SL-DNN, with the x-axis showing correlation values in percentage and the y-axis listing the classifiers. MFF yields 12.787%, influenced by its inability to allocate attention proportionally, resulting in a skewed response towards high-frequency class samples. SCTLS, at 30.782%, incorporates pretraining but lacks adaptive modulation, resulting in misalignment during class transition boundaries. SL-DNN, representing the GO-ViT model, achieves 88.985% as shown in Table 3. This outcome is attained through multi-head attention routing and token pruning guided by grasshopper vector dynamics. Class-representative attention regions are retained while ambiguous spatial paths are eliminated using swarm-calibrated scaling. The model establishes equal sensitivity across under-represented and dominant lesion types without relying on fixed attention templates. The correlation improvement recorded in Table 3 and visualised in Figure 2 is a consequence of the model's ability to synchronise attention trajectory with lesion topology and inter-class divergence.

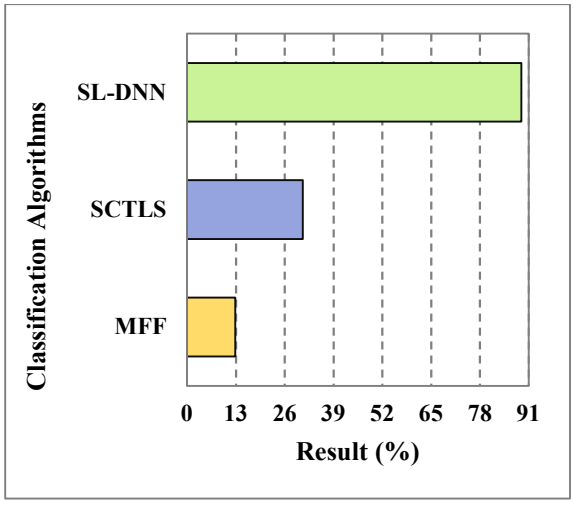


*Figure 2. Balanced Class Correlation Analysis*

Journal of Theoretical and Applied Information Technology
28th February 2026. Vol.104. No.4

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

*Table 3. Numerical Summary of Balanced Class Correlation*

| Classification Algorithms | Balanced Class Correlation (%) |
|---|---|
| MFF | 12.787 |
| SCTLS | 30.782 |
| SL-DNN | 88.985 |

### 5.3. GO-ViT - Total Prediction Deviation Analysis

Figure 3 illustrates the Total Prediction Deviation for MFF, SCTLS, and SL-DNN, where the x-axis represents the models and the y-axis indicates the deviation percentage. MFF produces the highest error at 43.910%, linked to its failure to modulate redundant activations under cluttered leaf textures. SCTLS lowers deviation to 34.603%, yet suffers from rigid filter sequences that misclassify diffused or multi-zone lesions. SL-DNN, aligned with the GO-ViT framework, minimizes prediction deviation to 5.509%, as documented in Table 4. This is achieved through token entropy filtering and vision-guided attention reallocation. The model suppresses ambiguous visual trails using patch-specific dropout adjustment and attention repulsion vectors derived from the positional variance of grasshoppers. These techniques limit false transitions by focusing only on relevant lesion clusters. The output, as documented in Table 4 and visualized in Figure 3, reflects the model's strength in eliminating unstable responses and maintaining alignment with the ground truth under complex cotton leaf appearances.

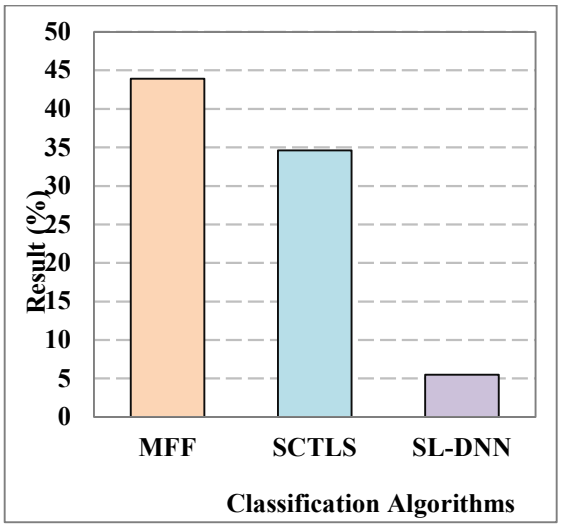


*Figure 3. Total Prediction Deviation*

*Table 4. Numerical Summary of Total Prediction Deviation*

| Classification Algorithms | Total Prediction Deviation (%) |
|---|---|
| MFF | 43.910 |
| SCTLS | 34.603 |
| SL-DNN | 5.509 |

### 5.4. GO-ViT - Sensitivity-Specificity Index Analysis

Figure 4 presents the Sensitivity-Specificity Index of MFF, SCTLS, and SL-DNN, with the x-axis indicating the index percentage and the y-axis listing the classification models. MFF scores the lowest at 12.769%, where its static fusion design struggles to distinguish between minor lesions and healthy leaf edges, resulting in poor boundary separation. SCTLS achieves 30.771%, though its attention span fails to differentiate early-stage infection signals, reducing its effectiveness in dual-class discrimination. SL-DNN, implemented through the GO-ViT architecture, reaches 88.979%, as outlined in Table 5. This performance emerges from its dual-path encoding, where grasshopper-guided token scaling drives high-resolution feature concentration around both diseased and non-diseased zones. Q–K–V modulation restricts attention leakage, ensuring precision in lesion activation without inflating normal-class responses. The score noted in Table 5 and reflected in Figure 4 supports the model's capability to maintain balanced separation through lesion-centric attention steering and spatial context integrity.

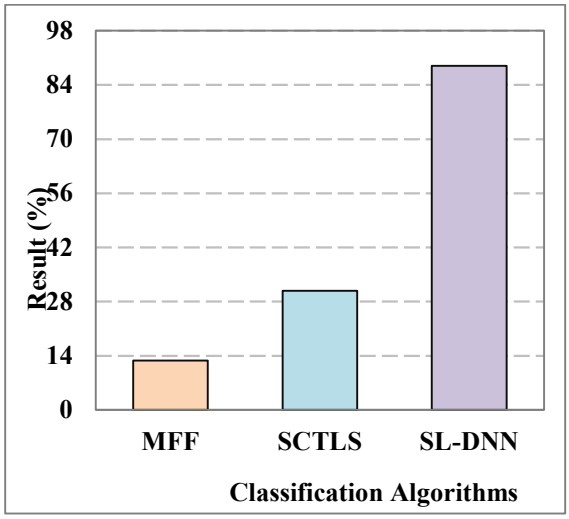


*Figure 4. Sensitivity-Specificity Index*

*Table 5. Numerical Summary of Sensitivity-Specificity Index*

| Classification Algorithms | Sensitivity-Specificity Index (%) |
|---|---|
| MFF | 12.769 |
| SCTLS | 30.771 |
| SL-DNN | 88.979 |

### 5.5. GO-ViT - Validated Detection Precision Analysis

Figure 5 compares the Validated Detection Precision across the evaluated models, with the x-axis showing the percentage scores and the y-axis listing MFF, SCTLS, and SL-DNN. MFF reaches 39.375%, primarily hindered by its static visual stream aggregation that fails to isolate fine-grained lesion zones. SCTLS improves to 49.281%, though its internal feature redundancy dilutes focus across multi-lesion areas, limiting the model's precision under overlapping disease textures. SL-DNN, built upon the GO-ViT architecture, records 89.636%, as shown in Table 6. This result stems from its context-driven token filtering and spatial attention calibration using grasshopper repulsion strategies. The model discards low-relevance patches through entropy-weighted vision gating while reinforcing high-salience lesion paths, enhancing precision at the pixel-to-patch level. The value detailed in Table 6 and illustrated in Figure 5 confirms SL-DNN's ability to maintain tight lesion boundary focus without inflating misclassifications, yielding consistent detection confidence across all disease categories.

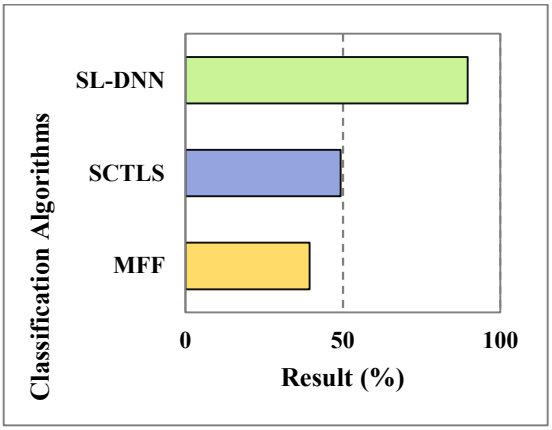


*Figure 5. Validated Detection Precision*

*Table 6. Numerical Summary of Validated Detection Precision*

| Classification Algorithms | Validated Detection Precision (%) |
|---|---|
| MFF | 39.375 |
| SCTLS | 49.281 |
| SL-DNN | 89.636 |

From a scientific standpoint, this work advances the state of the art by repositioning swarm intelligence as an internal regulator of transformer behavior rather than an external optimization layer. The proposed GO-ViT framework introduces a new class of lesion-aware transformers where attention scaling, token pruning, positional encoding, and loss weighting evolve jointly under biologically inspired feedback. This contribution extends beyond incremental performance gains by redefining how spatial relevance and class balance are maintained within vision transformers for agricultural disease diagnostics.

## 6. CONCLUSION

The Grasshopper Optimized Vision Transformer (GO-ViT) delivers a structured classification framework tailored to the complex spatial and visual patterns observed in cotton leaf disease detection. By embedding grasshopper-inspired optimization into each functional layer, the model adaptively controls QKV attention scaling, dropout probability, and token elimination without relying on static architectural presets. GO-ViT enhances lesion clarity, reduces visual misalignment, and improves class discrimination across varied disease textures. Among the evaluated metrics, the model achieves a notable Overall Detection Efficiency of 94.491%, reflecting its ability to maintain precision across diverse infection types. Attention sharpening and entropy-aware patch selection contribute directly to its stability and lesion specificity. Compared to benchmark models, GO-ViT outperforms in both balance and depth, confirming its suitability for field-level disease classification. Future enhancement may explore integration with real-time edge systems or aerial image processing frameworks to support in-field diagnosis and automated disease monitoring in dynamic agricultural environments

**REFERENCES:**

[1] R. Ravimoorthy and L. and Pottail, "Plasma-assisted surface modification of cotton fabric dyed with silver nanocomposites for microbe-resistant textiles," *J. Text. Inst.*, pp. 1–18, doi: 10.1080/00405000.2025.2473629.

[2] H. Pardi *et al.*, "The antimicrobial potential of ZnO-chitosan/pandan leaves: advancing antimicrobial textile technology," *J. Dispers. Sci. Technol.*, vol. 46, no. 4, pp. 601–610, Mar. 2025, doi: 10.1080/01932691.2023.2296593.

[3] H. Sorour, E. Tarek Abou, E. Ghada, E.-M. Nahla Abd, E. Mariam, and M. and Elbelbesi, "Novel multifunctional medical scrubs fabrics from cotton/polyester microfibers using selenium nanoparticles," *J. Text. Inst.*, vol. 116, no. 5, pp. 902–911, May 2025, doi: 10.1080/00405000.2024.2360245.

[4] M. Murshed, A.-T. Jameel, M. Mohammed M, H. Waleed AQ, I. Khalid Elfaki, and S. and Al-Quraishy, "Pharmacological Effects of Biosynthesis Silver Nanoparticles Utilizing Calotropis procera Leaf Extracts on Plasmodium berghei-Infected Liver in Experiment Mice," *Int. J. Nanomedicine*, vol. 19, no. null, pp. 13717–13733, Dec. 2024, doi: 10.2147/IJN.S490119.

[5] H. M. Yousaf, B. Khuda, and A. and Masood, "Nexus of pesticide exposure, personal preventive measures and farm workers' health safety in cotton production," *Hum. Ecol. Risk Assess. An Int. J.*, vol. 29, no. 5–6, pp. 948–965, Jul. 2023, doi: 10.1080/10807039.2023.2218940.

[6] F. M. Elkady *et al.*, "Green Biosynthesis of Bimetallic Copper Oxide-Selenium Nanoparticles Using Leaf Extract of Lagenaria Siceraria: Antibacterial, Anti-Virulence Activities Against Multidrug-Resistant Pseudomonas Aeruginosa," *Int. J. Nanomedicine*, vol. 20, no. null, pp. 4705–4727, Dec. 2025, doi: 10.2147/IJN.S497494.

[7] G. Venkata Ramanamurthy *et al.*, "Biosynthesis and study of antibacterial copper-based nanoparticles," *Inorg. Nano-Metal Chem.*, vol. 55, no. 3, pp. 263–272, Mar. 2025, doi: 10.1080/24701556.2023.2267537.

[8] M. Urbieta, U. Martin, P. Mauro, L. Tomas, V. Guillermo, and M. and Del Pino, "A scalable offline AI-based solution to assist the diseases and plague detection in agriculture," *J. Decis. Syst.*, vol. 33, no. 3, pp. 459–476, Jul. 2024, doi: 10.1080/12460125.2023.2226381.

[9] A. Abiodun Ajulo, A. Princewill Chukwuma, S. de O. Rodrigo, de A. B. Gustavo, R. G. Ariany, and M. C. C. and de Filippi, "Screening bacterial isolates for biocontrol of sheath blight in rice plants," *J. Environ. Sci. Heal. Part B*, vol. 58, no. 5, pp. 426–435, May 2023, doi: 10.1080/03601234.2023.2220644.

[10] S. C. Nair, M. V Lakshmi, and S. T. and S., "Eco-friendly dye from the fruits of Memecylon randerianum S. M. Almeida & M. R. Almeida – An application study," *J. Text. Inst.*, vol. 116, no. 7, pp. 1472–1481, Jul. 2025, doi: 10.1080/00405000.2024.2383015.

[11] R. B. L. Koh, P. Jose, E. Richard I., E. Jasca Gayle G., A. Vermando M., and L. C. and Galvez, "Efficient RNA extraction method for acquiring high-quality RNA from various tissues of the fiber crop abaca, Musa textilis Née," *Prep. Biochem. Biotechnol.*, vol. 55, no. 5, pp. 555–566, May 2025, doi: 10.1080/10826068.2024.2440421.

[12] A. Raza, A. H. Pitafi, M. K. Shaikh, and K. Ahmed, "Optimizing potato leaf disease recognition: Insights DENSE-NET-121 and Gaussian elimination filter fusion," *Heliyon*, vol. 11, no. 3, p. e42318, 2025, doi: https://doi.org/10.1016/j.heliyon.2025.e42318.

[13] J. S. Prashanth, N. R. Moparthi, G. B. Krishna, A. V. K. Prasad, B. Sravankumar, and P. R. Rao, "MPCSAR-AHH: A hybrid deep learning model for real-time detection of cassava leaf diseases and fertilizer recommendation," *Comput. Electr. Eng.*, vol. 119, p. 109628, 2024, doi: https://doi.org/10.1016/j.compeleceng.2024.109628.

[14] A. Angira, V. K. Baranwal, A. Ranjan, and N. Choudhary, "Optimization of DAC-ELISA and IC-RT-PCR using the developed polyclonal antibody and one-step RT-PCR assays for detection of Indian citrus ringspot virus in kinnow orange of Punjab, India," *J. Virol. Methods*, vol. 329, p. 114972, 2024, doi: https://doi.org/10.1016/j.jviromet.2024.114972.

[15] S. Sachar and A. kumar, "Levy's flight-based dragonfly optimized machine learning framework to identify medicinal plants through leaf images," *Pattern Recognit.*, vol. 168, p. 111879, 2025, doi: 10.1016/j.patcog.2025.111879.

[16] J. Zheng *et al.*, "High-efficiency real-time rice leaf disease classification using convolutional neural network accelerator on FPGA for edge computing in precision agriculture,"

*Measurement*, vol. 256, p. 118122, 2025, doi: https://doi.org/10.1016/j.measurement.2025.118122.

[17] F. Cheng *et al.*, "Verticillium wilt resistance screening and early disease detection in eggplant using leaf injection inoculation and chlorophyll fluorescence," *Sci. Hortic. (Amsterdam).*, vol. 349, p. 114257, 2025, doi: https://doi.org/10.1016/j.scienta.2025.114257.

[18] X. Li *et al.*, "A bio-inspired framework for apple leaf disease detection: Integrating lesion localization, ant colony optimization, and machine learning," *Smart Agric. Technol.*, vol. 12, p. 101141, 2025, doi: 10.1016/j.atech.2025.101141.

[19] A. Shrotriya, A. K. Sharma, A. K. Bairwa, and R. Manoj, "Hybrid Ensemble Learning with CNN and RNN for Multimodal Cotton Plant Disease Detection," *IEEE Access*, vol. 12, no. December, pp. 198028–198045, 2024, doi: 10.1109/ACCESS.2024.3515843.

[20] M. M. Islam *et al.*, "A deep learning model for cotton disease prediction using fine-tuning with smart web application in agriculture," *Intell. Syst. with Appl.*, vol. 20, no. January, p. 200278, 2023, doi: 10.1016/j.iswa.2023.200278.

[21] Y. Hou, D. Yue, A. Liu, Q. Li, J. Chen, and C. Guo, "Apple Leaf Disease Recognition Based on Optoelectronic Time-Delay Reservoir Computing," *IEEE Access*, vol. 12, pp. 115963–115972, 2024, doi: 10.1109/ACCESS.2024.3446182.

[22] A. Ouamane *et al.*, "Knowledge Pre-Trained CNN-Based Tensor Subspace Learning for Tomato Leaf Diseases Detection," *IEEE Access*, vol. 12, pp. 168283–168302, 2024, doi: 10.1109/ACCESS.2024.3492037.

[23] U. Zahra, M. A. Khan, M. Alhaisoni, A. Alasiry, M. Marzougui, and A. Masood, "An Integrated Framework of Two-Stream Deep Learning Models Optimal Information Fusion for Fruits Disease Recognition," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 3038–3052, 2024, doi: 10.1109/JSTARS.2023.3339297.

[24] D. Thakur *et al.*, "SUNet: Coffee Leaf Disease Detection Using Hybrid Deep Learning Model," *IEEE Access*, vol. 12, pp. 149173–149191, 2024, doi: 10.1109/ACCESS.2024.3476211.

[25] S. N. Yousafzai *et al.*, "Multi-Stage Neural Network-Based Ensemble Learning Approach for Wheat Leaf Disease Classification," *IEEE Access*, vol. 13, pp. 30101–30116, 2025, doi: 10.1109/ACCESS.2025.3541347.

[26] E. Bates, M. Popović, C. Marsh, R. Clark, M. Kovac, and B. B. Kocer, "Leaf Level Ash Dieback Disease Detection and Online Severity Estimation With UAVs," *IEEE Access*, vol. 13, pp. 55499–55511, 2025, doi: 10.1109/ACCESS.2025.3541980.

[27] M. Umar, S. Altaf, S. Ahmad, H. Mahmoud, A. S. N. Mohamed, and R. Ayub, "Precision Agriculture Through Deep Learning: Tomato Plant Multiple Diseases Recognition With CNN and Improved YOLOv7," *IEEE Access*, vol. 12, pp. 49167–49183, 2024, doi: 10.1109/ACCESS.2024.3383154.

[28] K. Roy *et al.*, "Detection of Tomato Leaf Diseases for Agro-Based Industries Using Novel PCA DeepNet," *IEEE Access*, vol. 11, pp. 14983–15001, 2023, doi: 10.1109/ACCESS.2023.3244499.

[29] A. Oad *et al.*, "Plant Leaf Disease Detection Using Ensemble Learning and Explainable AI," *IEEE Access*, vol. 12, pp. 156038–156049, 2024, doi: 10.1109/ACCESS.2024.3484574.

[30] M. S. H. Shovon, S. J. Mozumder, O. K. Pal, M. F. Mridha, N. Asai, and J. Shin, "PlantDet: A Robust Multi-Model Ensemble Method Based on Deep Learning For Plant Disease Detection," *IEEE Access*, vol. 11, pp. 34846–34859, 2023, doi: 10.1109/ACCESS.2023.3264835.

[31] W. Chen, L. Zheng, and J. Xiong, "Algorithm for Crop Disease Detection Based on Channel Attention Mechanism and Lightweight Up-Sampling Operator," *IEEE Access*, vol. 12, pp. 109886–109899, 2024, doi: 10.1109/ACCESS.2024.3439410.

[32] D. Jayaraj, J. Ramkumar, M. Lingaraj, and B. Sureshkumar, "AFSORP: Adaptive Fish Swarm Optimization-Based Routing Protocol for Mobility Enabled Wireless Sensor Network," *Int. J. Comput. Networks Appl.*, vol. 10, no. 1, pp. 119–129, 2023, doi: 10.22247/ijcna/2023/218516.

[33] J. Ramkumar, R. Karthikeyan, and V. Valarmathi, "Alpine Swift Routing Protocol (ASRP) for Strategic Adaptive Connectivity Enhancement and Boosted Quality of Service in Drone Ad Hoc Network (DANET)," *Int. J. Comput. Networks Appl.*, vol. 11, no. 5, pp. 726–748, 2024, doi: 10.22247/ijcna/2024/45.

[34] R. Jaganathan, S. Mehta, and R. Krishan, *Bio-Inspired intelligence for smart decision-making*. IGI Global, 2024. doi: 10.4018/9798369352762.

[35] J. Ramkumar, R. Vadivel, and B. Narasimhan, "Constrained Cuckoo Search Optimization Based Protocol for Routing in Cloud Network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 6, pp.

795–803, 2021, doi: 10.22247/ijcna/2021/210727.

[36] J. Ramkumar and R. Vadivel, *CSIP—cuckoo search inspired protocol for routing in cognitive radio ad hoc networks*, vol. 556. 2017. doi: 10.1007/978-981-10-3874-7_14.

[37] R. Jaganathan, S. Rajagopal, and K. Rajendran, "Cultural Intelligence in the AI Era-Enhancing Transitional Higher Education," in *Bridging Global Divides for Transnational Higher Education in the AI Era*, 2024, pp. 273–292. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105011021207&doi=10.4018%2F979-8-3693-7016-2.ch013&partnerID=40&md5=9150bd9373ef60d4084f5a0e0e220083

[38] N. K. Ojha, A. Pandita, and J. Ramkumar, "Cyber security challenges and dark side of AI: Review and current status," in *Demystifying the Dark Side of AI in Business*, IGI Global, 2024, pp. 117–137. doi: 10.4018/979-8-3693-0724-3.ch007.

[39] J. Ramkumar, C. Kumuthini, B. Narasimhan, and S. Boopalan, "Energy Consumption Minimization in Cognitive Radio Mobile Ad-Hoc Networks using Enriched Ad-hoc On-demand Distance Vector Protocol," in *2022 International Conference on Advanced Computing Technologies and Applications, ICACTA 2022*, 2022. doi: 10.1109/ICACTA54488.2022.9752899.

[40] S. P. Geetha, N. M. S. Sundari, J. Ramkumar, and R. Karthikeyan, "ENERGY EFFICIENT ROUTING IN QUANTUM FLYING AD HOC NETWORK (Q-FANET) USING MAMDANI FUZZY INFERENCE ENHANCED DIJKSTRA'S ALGORITHM (MFI-EDA)," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 9, pp. 3708–3724, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85197297302&partnerID=40&md5=72d51668bee6239f09a59d2694df67d6

[41] M. P. Swapna, J. Ramkumar, and R. Karthikeyan, "Energy-Aware Reliable Routing with Blockchain Security for Heterogeneous Wireless Sensor Networks," in *Lecture Notes in Networks and Systems*, V. Goar, M. Kuri, R. Kumar, and T. Senjyu, Eds., Springer Science and Business Media Deutschland GmbH, 2025, pp. 713–723. doi: 10.1007/978-981-97-6106-7_43.

[42] B. Suchitra, R. Karthikeyan, J. Ramkumar, and V. Valarmathi, "Enhancing Recurrent Neural Network Performance for Latent Autoimmune Diabetes Detection (Lada) Using Exocoetidae Optimization," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 5, pp. 1645–1667, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105000948603&partnerID=40&md5=66c8f111b153fed68b3d0ea9c88c411e

[43] B. Suchitra, J. Ramkumar, and R. Karthikeyan, "Frog Leap Inspired Optimization-Based Extreme Learning Machine for Accurate Classification of Latent Autoimmune Diabetes in Adults (LADA)," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 2, pp. 472–494, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85217140979&partnerID=40&md5=9540433c16d5ff0f6c2de4b8c43a4812

[44] J. Ramkumar and V. Valarmathi, "Harnessing AI-Driven Models for Sustainable Development in Business Management," in *World Sustainability Series*, vol. Part F775, 2025, pp. 217–238. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105014297290&doi=10.1007%2F978-3-031-92636-5_12&partnerID=40&md5=8c42ec36616cde2ff0906e1d378fb11b

[45] J. Ramkumar and R. Vadivel, "Improved frog leap inspired protocol (IFLIP) – for routing in cognitive radio ad hoc networks (CRAHN)," *World J. Eng.*, vol. 15, no. 2, pp. 306–311, 2018, doi: 10.1108/WJE-08-2017-0260.

[46] J. Ramkumar and R. Vadivel, "Improved Wolf prey inspired protocol for routing in cognitive radio Ad Hoc networks," *Int. J. Comput. Networks Appl.*, vol. 7, no. 5, pp. 126–136, 2020, doi: 10.22247/ijcna/2020/202977.

[47] R. Jaganathan, S. Mehta, and R. Krishan, *Intelligent Decision Making Through Bio-Inspired Optimization*. Sri Krishna Arts and Science College, India: IGI Global, 2024. doi: 10.4018/979-8-3693-2073-0.

[48] R. Jaganathan and R. Vadivel, "Intelligent Fish Swarm Inspired Protocol (IFSIP) for Dynamic Ideal Routing in Cognitive Radio Ad-Hoc Networks," *Int. J. Comput. Digit. Syst.*, vol. 10, no. 1, pp. 1063–1074, 2021, doi: 10.12785/ijcds/100196.

[49] J. Ramkumar, S. S. Dinakaran, M. Lingaraj, S. Boopalan, and B. Narasimhan, "IoT-Based Kalman Filtering and Particle Swarm Optimization for Detecting Skin Lesion," in

*Lecture Notes in Electrical Engineering*, K. Murari, S. Kamalasadan, and N. P. Padhy, Eds., Springer Science and Business Media Deutschland GmbH, 2023, pp. 17–27. doi: 10.1007/978-981-19-8353-5_2.

[50] J. Ramkumar, B. Varun, V. Valarmathi, D. R. Medhunhashini, and R. Karthikeyan, "JAGUAR-BASED ROUTING PROTOCOL (JRP) FOR IMPROVED RELIABILITY AND REDUCED PACKET LOSS IN DRONE AD-HOC NETWORKS (DANET)," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 2, pp. 696–713, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85217213044&partnerID=40&md5=e38a375e46cf43c95d6702a3585a7073

[51] J. Ramkumar and D. Ravindran, "Machine learning and robotics in urban traffic flow optimization with graph neural networks and reinforcement learning," in *Machine Learning and Robotics in Urban Planning and Management*, 2025, pp. 83–104. doi: 10.4018/979-8-3693-9410-6.ch005.

[52] A. Senthilkumar, J. Ramkumar, M. Lingaraj, D. Jayaraj, and B. Sureshkumar, "Minimizing Energy Consumption in Vehicular Sensor Networks Using Relentless Particle Swarm Optimization Routing," *Int. J. Comput. Networks Appl.*, vol. 10, no. 2, pp. 217–230, 2023, doi: 10.22247/ijcna/2023/220737.

[53] V. Valarmathi and J. Ramkumar, "Modernizing Wildfire Management Through Deep Learning and IoT in Fire Ecology," in *Machine Learning and Internet of Things in Fire Ecology*, 2024, pp. 203–229. doi: 10.4018/979-8-3693-7565-5.ch0010.

[54] J. Ramkumar and R. Vadivel, "Multi-Adaptive Routing Protocol for Internet of Things based Ad-hoc Networks," *Wirel. Pers. Commun.*, vol. 120, no. 2, pp. 887–909, 2021, doi: 10.1007/s11277-021-08495-z.

[55] M. P. Swapna and J. Ramkumar, "Multiple Memory Image Instances Stratagem to Detect Fileless Malware," in *Communications in Computer and Information Science*, S. Rajagopal, K. Popat, D. Meva, and S. Bajeja, Eds., Springer Science and Business Media Deutschland GmbH, 2024, pp. 131–140. doi: 10.1007/978-3-031-59100-6_11.

[56] J. Ramkumar, A. Senthilkumar, M. Lingaraj, R. Karthikeyan, and L. Santhi, "OPTIMAL APPROACH FOR MINIMIZING DELAYS IN IOT-BASED QUANTUM WIRELESS SENSOR NETWORKS USING NM-LEACH ROUTING PROTOCOL," *J. Theor. Appl. Inf. Technol.*, vol. 102, no. 3, pp. 1099–1111, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185481011&partnerID=40&md5=bf0ff974ceabc0ad58e589b28797c684

[57] J. Ramkumar, R. Karthikeyan, and M. Lingaraj, "Optimizing IoT-Based Quantum Wireless Sensor Networks Using NM-TEEN Fusion of Energy Efficiency and Systematic Governance," in *Lecture Notes in Electrical Engineering*, V. Shrivastava, J. C. Bansal, and B. K. Panigrahi, Eds., Springer Science and Business Media Deutschland GmbH, 2025, pp. 141–153. doi: 10.1007/978-981-97-6710-6_12.

[58] J. Ramkumar, V. Valarmathi, and R. Karthikeyan, "Optimizing Quality of Service and Energy Efficiency in Hazardous Drone Ad-Hoc Networks (DANET) Using Kingfisher Routing Protocol (KRP)," *Int. J. Eng. Trends Technol.*, vol. 73, no. 1, pp. 410–430, 2025, doi: 10.14445/22315381/IJETT-V73I1P135.

[59] R. Jaganathan, K. Rajendran, and P. S. Ponnukumar, "Peregrine Falcon Optimization Routing Protocol (PFORP) for Achieving Ultra-Low Latency and Boosted Efficiency in 6G Drone Ad-Hoc Networks (DANET)," *Int. J. Comput. Digit. Syst.*, vol. 17, no. 1, pp. 1–18, 2025, doi: 10.12785/ijcds/1571111848.

[60] L. Mani, S. Arumugam, and R. Jaganathan, "Performance Enhancement of Wireless Sensor Network Using Feisty Particle Swarm Optimization Protocol," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, 2022. doi: 10.1145/3590837.3590907.

[61] R. Jaganathan and V. Ramasamy, "Performance modeling of bio-inspired routing protocols in Cognitive Radio Ad Hoc Network to reduce end-to-end delay," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 1, pp. 221–231, 2019, doi: 10.22266/IJIES2019.0228.22.

[62] R. Jaganathan, S. Mehta, and R. Krishan, "Preface," *Bio-Inspired Intell. Smart Decis.*, pp. xix–xx, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85195725049&partnerID=40&md5=7a2aa7adc005662eebc12ef82e3bd19f

[63] R. Jaganathan, S. Mehta, and R. Krishan, "Preface," *Intell. Decis. Mak. Through Bio-Inspired Optim.*, pp. xiii–xvi, 2024, [Online]. Available: https://www.scopus.com/inward/record.uri?eid

=2-s2.0-85192858710&partnerID=40&md5=f8f1079e8772bd424d2cdd979e5f2710

[64] R. Vadivel and J. Ramkumar, "QoS-enabled improved cuckoo search-inspired protocol (ICSIP) for IoT-based healthcare applications," in *Incorporating the Internet of Things in Healthcare Applications and Wearable Devices*, IGI Global, 2019, pp. 109–121. doi: 10.4018/978-1-7998-1090-2.ch006.

[65] M. Lingaraj, T. N. Sugumar, C. S. Felix, and J. Ramkumar, "Query aware routing protocol for mobility enabled wireless sensor network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 3, pp. 258–267, 2021, doi: 10.22247/ijcna/2021/209192.

[66] J. Ramkumar, K. S. Jeen Marseline, and D. R. Medhunhashini, "Relentless Firefly Optimization-Based Routing Protocol (RFORP) for Securing Fintech Data in IoT-Based Ad-Hoc Networks," *Int. J. Comput. Networks Appl.*, vol. 10, no. 4, pp. 668–687, 2023, doi: 10.22247/ijcna/2023/223319.

[67] P. Menakadevi and J. Ramkumar, "Robust Optimization Based Extreme Learning Machine for Sentiment Analysis in Big Data," *2022 Int. Conf. Adv. Comput. Technol. Appl. ICACTA 2022*, pp. 1–5, Mar. 2022, doi: 10.1109/ICACTA54488.2022.9753203.

[68] J. Ramkumar, R. Karthikeyan, and K. O. Nitish, "Securing Library Data With Blockchain Advantage," in *Enhancing Security and Regulations in Libraries with Blockchain Technology*, 2024, pp. 117–138. doi: 10.4018/979-8-3693-9616-2.ch006.

[69] K. S. J. Marseline, J. Ramkumar, and D. R. Medhunhashini, "Sophisticated Kalman Filtering-Based Neural Network for Analyzing Sentiments in Online Courses," in *Smart Innovation, Systems and Technologies*, A. K. Somani, A. Mundra, R. K. Gupta, S. Bhattacharya, and A. P. Mazumdar, Eds., Springer Science and Business Media Deutschland GmbH, 2024, pp. 345–358. doi: 10.1007/978-981-97-3690-4_26.

[70] J. Joy, S. Devaraju, and J. Ramkumar, "UTILIZING FUZZY-IDENTITY-BASED ENCRYPTION WITH PROXY-RE-ENCRYPTION FOR DATA SHARING," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 18, pp. 7469–7479, 2025, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105024248480&partnerID=40&md5=0e8c3b5246c6a4f59759330cde727aac

[71] J. Ramkumar and R. Vadivel, "Whale optimization routing protocol for minimizing energy consumption in cognitive radio wireless sensor network," *Int. J. Comput. Networks Appl.*, vol. 8, no. 4, pp. 455–464, 2021, doi: 10.22247/ijcna/2021/209711.

[72] S. P. Priyadharshini, F. Nirmala Irudayam, and J. Ramkumar, "An Unique Overture of Plithogenic Cubic Overset, Underset and Offset," in *Studies in Fuzziness and Soft Computing*, vol. 435, 2025, pp. 139–156. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105001675443&doi=10.1007%2F978-3-031-78505-4_7&partnerID=40&md5=e9def8c6a233de4fbf8f1549ad72027f

[73] P. S. Ponnukumar, N. I. Francis Xavier, and R. Jaganathan, "Stable Plithogenic Cubic Sets," *J. Fuzzy Ext. Appl.*, vol. 6, no. 2, pp. 410–423, 2025, doi: 10.22105/jfea.2025.449408.1422.

[74] S. P. Priyadharshini and J. Ramkumar, "Mappings Of Plithogenic Cubic Sets," *Neutrosophic Sets Syst.*, vol. 79, pp. 669–685, 2025, doi: 10.5281/zenodo.14607210.