Hybrid CNN–Transformer Model for Maize Leaf Blight Classification using Adaptive Genetic Optimization

Akhilesh Kumar
Dr. Ram Manohar Lohia Avadh University
Lokendra Singh Umrao
lsucs@mmmut.ac.in
Madan Mohan Malaviya University of Technology

Research Article

Keywords: Maize leaf blight detection, Vision Transformer, Convolutional Neural Network, Adaptive Genetic Optimization, Deep learning, Crop disease classification

Posted Date: February 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8783155/v1

Additional Declarations: No competing interests reported.

Hybrid CNN–Transformer Model for Maize Leaf Blight Classification using Adaptive Genetic Optimization

Akhilesh Kumar[1] and Lokendra Singh Umrao[2,*]

[1]Department of Computer Science and Engineering, Dr. Ram Manohar Lohia Avadh University, Ayodhya, India

Email: kumar.aki297@gmail.com

[2]Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, India

[*]Corresponding author. E-mail: lsucs@mmmut.ac.in

Abstract

Maize leaf blight is a disastrous foliar disease in the world production of maize that causes significant losses in terms of yield annually. The classical machine learning and convolutional neural network (CNN) models are prone to poor generalization across the different conditions in the field because of differences in lighting, background, and leaf morphology. To counter these difficulties, this research suggests the use of Hybrid CNN-Transformer architecture that is trained using the Adaptive Genetic Optimization (AGO) to provide accurate classification of maize leaf blight. The hybrid model will utilize the spatial feature extraction and Global attention mechanism of Vision Transformer (ViT) to extract local and contextual patterns of diseases. AGO algorithm is dynamically adjusted with essential hyperparameters, such as the learning rate, filter size, and embedding, and adjusted according to the population diversity and fitness assessment, thus enhancing the convergence speed and classification accuracy. The experimental analysis of an augmented dataset of maize leaf disease proves that the developed model has a higher classification accuracy of 98.1, compared to the base models, including ResNet50 (96.7) and MobileNetV3 (97.2). The hybrid AGOCNN Transformer version demonstrates better solutions to the intelligent system on agricultural disease management in the real-field conditions.

Keywords

Maize leaf blight detection, Vision Transformer, Convolutional Neural Network, Adaptive Genetic Optimization, Deep learning, Crop disease classification

1. Introduction

Maize (Zea mays L.) is the most important cereal crop worldwide with its application as staple food and livestock feed, as well as a raw material in the industry in the production of biofuels and starch (Ahila Priyadharshini et al., 2019). Maize productivity is however under serious attack due to various foliar diseases but one of the most destructive is the maize leaf blight, which is caused by Helminthosporium turcicum (Borhani et al., 2022). The illness presents as long necrotic spots on the foliage, which gradually decrease the power of photosynthesis and cause loss of yield up to 50 percent in suitable circumstances. Therefore, it is important to identify maize leaf blight at an early stage and correctly to enhance crop management, lessen reliance on chemical pesticides, and maintain agricultural productivity (Thakur et al., 2023).

The conventional methods of diagnosis of leaf diseases have been based on manual examination under the eye of the agricultural expert which is both time consuming, subjective and subject to human error (Waheed et al., 2020). In automating this process, support vector machines (SVM), random forest (RF), and K-means clustering which are classical methods of machine learning have been used to analyze the leaf images. Such techniques rely on hand-crafted feature extraction, i.e. color, texture and shape descriptions are hand-crafted. Nevertheless, this type of handcrafted properties is very susceptible to environmental factors such as the difference in light, background distortion, and occlusion leading to poor robustness and scalability when used in real field (Qian et al., 2022).

With the availability of Deep Learning (DL) and Convolutional Neural Networks (CNNs), the process of plant disease detection has been greatly advanced due to the automatic hierarchical extraction of features. CNNs are useful in capturing local texture patterns and color variations of surfaces of the leaves. The small receptive field however limits the model to discern contextual dependencies in the world over spatial areas, which frequently lead to misclassification in situations where similar local features occur in disparate disease groups (Li & Li, 2022).

Vision Transformers (ViT) have appeared in order to overcome these difficulties as a strong alternative that can capture long-range dependencies with the help of self-attention mechanisms. ViTs are better at capturing inter-patch relationships and are better contextually understood than CNNs. Nonetheless, their effectiveness is greatly inspired by many hyperparameters, including patch size, embedding dimensions, attention heads, and learning rates. A poor adjustment of these parameters may result in overfitting, poor convergence and poor accuracy (Arshad et al., 2023).

This study presents a Hybrid CNN -Transformer Model that uses Adaptive Genetic Optimization (AGO) to detect maize leaf blight. The hybrid model is a hybrid of CNN and the Transformer, which extracts local features and renders the global contextual representation, respectively, and AGO optimizes the hyper parameters of the model automatically by dynamically conducting evolutionary search (Taji et al., 2024). This combination can improve the efficiency of learning, speed, and convergence, as well as provide high Classification accuracy across variable field settings. The framework proposed has better robustness, scalability and generalization compared to current models in the state-of-the-art (Dosovitskiy et al., 2020).

2. Related Work

Deep learning in identifying plant diseases has been of great interest within recent years, especially maize leaf diseases. Initial methods used were mainly based on classic image processing and shallow learning schemes that were later replaced by convolutional neural network (CNN)-based systems because they had better learnings on features.

One of the earliest deep learning-based maize disease classifiers, (Ahila Priyadharshini et al., 2019) used deep convolutional neural networks (DCNNs) in their experiments, demonstrating better results than the traditional handcrafted feature extraction models. Likewise, (Waheed et al., 2020) came up with an optimized Dense CNN to recognize corn leaf disease, which stated that more complex architectures can be more robust, but a precise tuning of hyperparameters is needed to converge and achieve high accuracy. (EL Da Rocha et al., 2021) also noted that hyperparameter optimization is significant to enhance CNN performance, yet optimization is still expensive computationally.

Recent developments have delved into attention mechanisms and light weight networks to augment the efficiency in maize disease detection. (Chen et al., 2021) proposed an attention-based lightweight CNN that enhanced the recognition of maize diseases in changing fairly field conditions. (Qian et al., 2022) went further to add self-attention mechanisms to enhance feature representation in deep networks leading to a great improvement in detection performance in complicated backgrounds. Likewise, (Cui et al., 2023) designed a CBAM enhanced lightweight autoencoder to classify maize leaf disease with high accuracy and less model complexity.

Vision Transformers (ViT) have changed the face of image-based recognition by modelling the over long distance dependency with self-attention (Dosovitskiy et al., 2020). (Borhani et al., 2022) used ViT to classify plant diseases in an automated manner and showed that the model is effective in the process of representing global spatial relationships. (Pacal and Işik, 2024) integrated CNN and ViT framework to identify corn leaf disease with high accuracy whereas (Pacal, 2024) augmented ViT framework to determine maize disease in large-scale datasets with high predictability across different disease severity. (Ramadan et al., 2024) and (Wang et al., 2024) engaged in comparative studies of the CNN and ViT-based classifiers and determined that ViTs are more accurate than CNNs but have disadvantages in terms of computational cost and hyperparameter sensitivity.

A hybrid model that incorporates CNN and Transformer is found to be considered based on the capability to strike a balance between local and global contextual modeling. (Li and Li, 2022) suggested a lightweight hybrid CNN100100 Transformer to classify apple diseases and indicated that hybridization enhances the interpretability and accuracy. The synergy was confirmed by (Thakur et al., 2023), who used CNN feature extraction, but Transformer based self-attention to classify plant diseases. On the same note, (Arshad et al., 2023) introduced PLDPNet, an end-to-end deep learning model to predict potato diseases with superior performance in terms of end-to-end application of CNN and attention layers.

Similar studies have also investigated the metaheuristic optimization to improve the performance of deep models. (Taji et al., 2024) compared various optimization algorithms, such as Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) in terms of hybrid CNN-based detection of plant diseases, and determined that Genetic Algorithm (GA) is one of the best methods to accelerate convergence. Applications of Cat Swarm Optimization to maize leaf disease detection had shown competitive results by (Vimalkumar and Latha, 2024), and had shown that optimized SVMs with DenseNet201 features could greatly increase the accuracy of maize disease classification (Dash et al., 2023).

Despite these advancements, existing studies face limitations such as fixed hyperparameter settings, slow convergence, and reduced adaptability across diverse environmental conditions. The current work addresses these challenges through a Hybrid CNN–Transformer model optimized using Adaptive Genetic Optimization (AGO). The proposed approach introduces adaptive mutation and crossover probabilities based on population variance, ensuring dynamic exploration of the parameter space. This hybrid strategy effectively combines CNN’s local texture learning, Transformer's contextual awareness, and AGO's adaptive optimization, resulting in enhanced performance stability and classification precision compared to prior deep learning and metaheuristic models.

3. Proposed Methodology

The suggested architecture, which it will be called Hybrid CNN transformer with Adaptive Genetic Optimization (AGO), combines the high spatial representation of Convolutional Neural Networks (CNNs) and the contextual reasoning of Vision Transformers (ViT). This hybrid model is also improved by an Adaptive Genetic Optimization process which is continually used to optimize hyperparameters in order to obtain optimal convergence, accuracy and generalization.

The general scheme of the proposed system consists of four major steps: (1) image preprocessing and augmentation, (2) hybrid feature extraction through CNN-Transformer fusion, (3) adaptive genetic optimization to hyperparameter optimization, and (4) final disease classification.

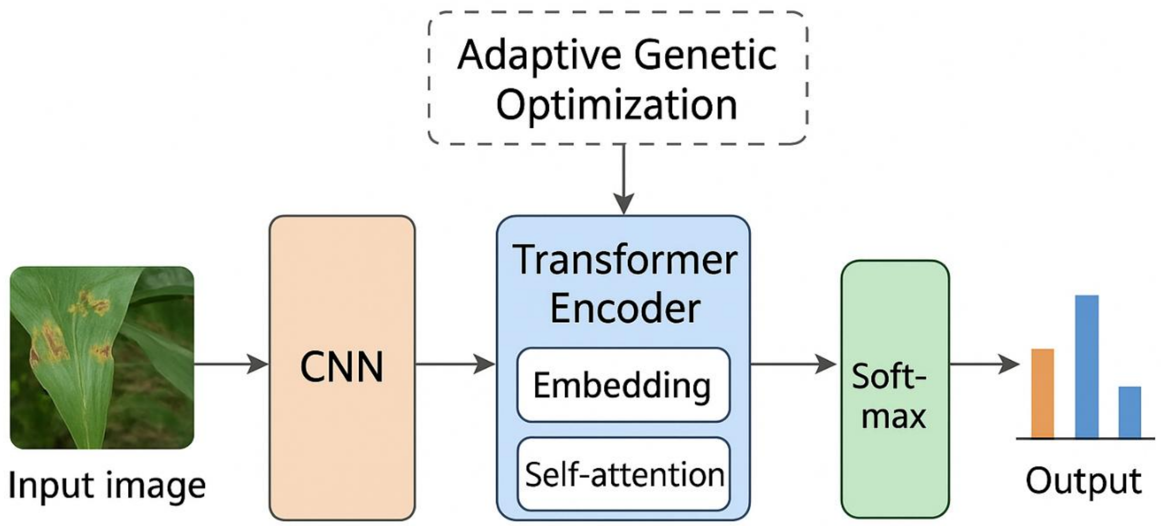

Figure 1: Proposed Model for Maize Leaf Blight Classification

System Architecture Overview

The complete architecture is depicted conceptually in Figure 1 (schematic representation). The model combines CNN's ability to learn localized texture features such as lesion edges, vein patterns, and discoloration with the Transformer's strength in capturing global contextual relationships across entire leaf surfaces.

- **Input Layer:** Preprocessed maize leaf images are fed into the CNN block for spatial feature extraction.
- **CNN Block:** Multiple convolutional and pooling layers generate low-level and mid-level feature maps representing texture and color distribution.
- **Feature Tokenization:** The CNN-generated feature maps are flattened into patch tokens, forming the input sequence for the Transformer encoder.
- **Transformer Encoder Block:** Multi-head self-attention captures inter-patch dependencies, ensuring contextual awareness of lesion distribution patterns.
- **Classification Head:** Aggregated features are passed through a dense layer and SoftMax classifier to determine the disease category.
- **Adaptive Genetic Optimizer:** A metaheuristic optimization algorithm adaptively adjusts key hyperparameters (e.g., learning rate, embedding dimension, and number of attention heads) to enhance training performance.

Image Preprocessing and Augmentation

Raw maize leaf images collected from open datasets and field surveys often suffer from lighting inconsistencies, occlusions, and background clutter. To ensure robustness and prevent overfitting, multiple preprocessing steps are applied:

- **Resizing and Normalization:** All images are resized to 224×224 pixels and normalized to the range [0, 1].
- **Histogram Equalization:** Enhances contrast and improves lesion visibility.
- **Data Augmentation:** Rotation (±15°), horizontal flipping, brightness adjustment, and Gaussian noise are applied to simulate real-field variations and enrich training diversity.

Mathematically, the normalized image $I'(x, y)$ is expressed as:

$$I'(x,y) = \frac{I(x,y) - \mu}{\sigma}$$

where μ and σ denote the mean and standard deviation of the image pixel intensities.

CNN Feature Extraction Module

The CNN module serves as the local feature extractor, learning spatial patterns corresponding to disease symptoms such as lesions, chlorosis, and necrosis. Each convolutional layer applies a kernel filter $K^{(l)}$ to the input feature map:

$$F_{i,j}^{(l)} = \sigma\left(\sum_{m,n} K_{m,n}^{(l)} \cdot I_{i+m,j+n'} + b^{(l)}\right)$$

where $\sigma(\cdot)$ denotes the ReLU activation function and $b^{(l)}$ represents the bias term. Max pooling layers are incorporated to reduce spatial dimensions while preserving discriminative features.

After the final convolutional stage, the feature map $F \in R^{w' \times h' \times c}$ is flattened into non-overlapping patchesP_k, where each patch corresponds to a local feature token that feeds into the Transformer encoder.

Vision Transformer Encoder Module

The Transformer encoder models global dependencies between spatial patches. Each patch token P_k is projected into an embedding vector Z_k using a trainable weight matrix W_e:

$$Z_k = W_e P_k + b_e$$

To preserve spatial order, positional encodings PE_kPEkare added to the embeddings:

$$Z_{k'} = Z_k + PE_k$$

Each embedding sequence undergoes multi-head self-attention (MHSA) computation, defined as:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)$$

where Q, K, V are the query, key, and value matrices derived from the linear transformations of $Z_{k'}$, and d_k is the dimension of each attention head.

The Transformer block output is then passed through a Feed Forward Network (FFN) and layer normalization to produce context-enriched feature representations:

$$O_i = \text{LayerNorm}\left(\text{FFN}\left(\text{MHA}(Z_{k'})\right)\right)$$

This stage allows the network to learn relationships between distant regions, thereby improving robustness to complex lesion distributions.

Adaptive Genetic Optimization (AGO)

To further improve training convergence and classification performance, an Adaptive Genetic Optimization strategy is integrated for dynamic hyperparameter tuning.

Initialization

A population of candidate solutions $P = \{\theta_1, \theta_2, \dots, \theta_n\}$ is initialized, where each individual θ_i represents a potential combination of hyperparameters such as:

$$\theta = \{lr,\ bs,\ d_{emb},\ h_{att}\}$$

Where lr = learning rate, bs = batch size, d_{emb} = embedding dimension, and h_{att} = number of attention heads.

Fitness Evaluation

Each individual's performance is measured using validation accuracy as the fitness function:

$$f_i = 1 - L(\theta_i)$$

Where $L(\theta_i)$ represents the model's loss function (categorical cross-entropy).

Adaptive Crossover and Mutation

The crossover (P_c) and mutation (P_m) probabilities are adjusted dynamically according to population diversity σ_f^2:

$$P_c = P_{c0}\left(1 - \frac{\sigma_f^2}{\sigma_{f,\max}^2}\right), P_m = P_{m0}\left(1 + \frac{\sigma_f^2}{\sigma_{f,\max}^2}\right)$$

This ensures a higher exploration rate when population variance decreases (preventing premature convergence) and greater exploitation when diversity is high (stabilizing learning).

Selection and Elitism

Top-performing individuals are selected through tournament selection, and elite solutions are carried forward to the next generation to preserve the best-found parameters.

Termination

The optimization continues until the maximum generation count $T_{\max}$ is reached or the improvement in fitness value Δf_i falls below a threshold ϵ.

Algorithmic Representation

Algorithm 1: Hybrid CNN–Transformer with Adaptive Genetic Optimization

Algorithm 1: Hybrid CNN–Transformer Model Optimized with Adaptive Genetic Optimization (AGO)
Input:
• $\boldsymbol{D}$: Maize leaf image dataset
• $\boldsymbol{P}$: Patch size
• $\boldsymbol{L}$: Number of Transformer layers
• $\boldsymbol{N}$: Number of attention heads
• $\boldsymbol{d}$: Embedding dimension
• $\boldsymbol{T_{max}}$: Maximum generations
• $\boldsymbol{\eta}$: Initial learning rate
• $\boldsymbol{P_{c0}}$: Initial crossover probability
• $\boldsymbol{P_{m0}}$: Initial mutation probability
• $\boldsymbol{\varepsilon}$: Convergence threshold
Output:
• Optimized parameter set $\boldsymbol{\theta}^* = \{\boldsymbol{lr}, \boldsymbol{bs}, \boldsymbol{d_{emb}}, \boldsymbol{h_{att}}, \boldsymbol{dr}\}$
Phase 1: Image Preprocessing
For each image $\boldsymbol{I} \in \boldsymbol{D}$:

1. Apply histogram equalization to enhance contrast.

2. Resize image to $\mathbf{224} \times \mathbf{224}$ pixels.

3. Normalize pixel values in range [0, 1].

4. Perform augmentation (rotation ±15°, flipping, noise injection).

5. Extract non-overlapping patches of size $\boldsymbol{P} \times \boldsymbol{P}$.

6. Flatten each patch into a 1D vector and store as a token.

7. Store preprocessed patches for CNN–Transformer input.

End For

Phase 2: CNN Feature Extraction and Patch Encoding

For each preprocessed image $\boldsymbol{I'}$:

8. Apply convolutional operations:

$$\boldsymbol{F_{i,j}^{(l)}} = \boldsymbol{\sigma}\left(\sum\nolimits_{\boldsymbol{m,n}} \boldsymbol{K_{m,n}^{(l)}} \cdot \boldsymbol{I_{i+m,j+n'}} + \boldsymbol{b^{(l)}}\right)$$

9. Generate spatial feature maps and flatten them into patch tokens.

10. For each token $\boldsymbol{P_k}$Pk, embed into a lower-dimensional space:

$$\boldsymbol{Z_k} = \boldsymbol{W_e} \cdot \text{Flatten}(\boldsymbol{P_k}) + \boldsymbol{b_e}$$

11. Add positional encodings to retain spatial order.

12. Collect all encoded tokens as input sequence $\boldsymbol{S}$S.

Phase 3: Vision Transformer Forward Pass

For each Transformer layer $\boldsymbol{l = 1}$

13. Apply Multi-Head Self-Attention to token sequence $\boldsymbol{S}$S:

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{V}$$

14. Pass attention output through Feed-Forward Network (FFN).

15. Normalize and update token embeddings.

16. Obtain final contextual feature representation $\boldsymbol{O}$O.

Apply SoftMax layer for classification:

$$\hat{\boldsymbol{y}} = \mathrm{Softmax}(\boldsymbol{W_c O} + \boldsymbol{b_c})$$

Phase 4: Adaptive Genetic Optimization (AGO)

Initialize:

- Population $\boldsymbol{P} = \{\boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \ldots, \boldsymbol{\theta_n}\}$ with random hyperparameter sets
- Generation counter $\boldsymbol{t} = \boldsymbol{0}$

While $\boldsymbol{t} < \boldsymbol{T_{max}}$

17. Fitness Evaluation:

- Train CNN–Transformer with parameters $\boldsymbol{\theta_i}$θi.
- Compute validation loss $\boldsymbol{L}(\boldsymbol{\theta_i})$ and fitness:

$$\boldsymbol{f_i} = \boldsymbol{1} - \boldsymbol{L}(\boldsymbol{\theta_i})$$

18. Selection:

- Sort individuals by descending $\boldsymbol{f_i}$fi.
- Retain top $\boldsymbol{k}$k as elite set.
- Select parents via tournament selection.

19. Diversity Calculation:

- Compute population variance:

$$\boldsymbol{\sigma_f^2} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{n}}\left(\boldsymbol{f_i} - \bar{\boldsymbol{f}}\right)^{\boldsymbol{2}}$$

20. Adaptive Control:

- Update crossover and mutation probabilities:

$$\boldsymbol{P_c} = \boldsymbol{P_{c0}}\left(\boldsymbol{1} - \frac{\boldsymbol{\sigma_f^2}}{\boldsymbol{\sigma_{f,\max}^2}}\right), \boldsymbol{P_m} = \boldsymbol{P_{m0}}\left(\boldsymbol{1} + \frac{\boldsymbol{\sigma_f^2}}{\boldsymbol{\sigma_{f,\max}^2}}\right)$$

21. Crossover:

- Generate offspring $\boldsymbol{\theta_{child}} = \boldsymbol{\alpha\theta_a} + (\boldsymbol{1} - \boldsymbol{\alpha})\boldsymbol{\theta_b}, \boldsymbol{\alpha} \in [\boldsymbol{0}, \boldsymbol{1}]$.

22. Mutation:

- Mutate parameters with probability $\boldsymbol{P_m}$Pm:

$$\boldsymbol{p_{j'}} = \boldsymbol{p_j} + \boldsymbol{\delta} \times \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{1})$$

23. Population Update:
○ Combine offspring and elite individuals to form new population $\boldsymbol{P'}$P′.
○ Replace $\boldsymbol{P} = \boldsymbol{P'}$
24. Increment $\boldsymbol{t} = \boldsymbol{t} + \mathbf{1}$
End While
Phase 5: Final Model Selection
25. Identify best-performing parameter set $\boldsymbol{\theta}^* = \mathbf{argmax}(\boldsymbol{f_i})$
26. Retrain Hybrid CNN–Transformer using $\boldsymbol{\theta}^*$ on full training + validation data.
27. Evaluate final model on test set using Accuracy, Precision, Recall, F1-score, and AUC.
Output: Optimized CNN–Transformer model with best parameter configuration $\boldsymbol{\theta}^*$θ∗ for real-time maize leaf blight detection.

The optimized CNN–Transformer outputs a set of contextual features OO, which are passed through a fully connected SoftMax classifier for disease prediction:

$$\hat{y} = \text{Softmax}(W_c O + b_c)$$

The output corresponds to one of the predefined maize disease categories, such as healthy, blight, or rust.

Advantages of the Proposed Approach

- **Hybrid Feature Learning:** Combines CNN’s localized spatial features with Transformer’s global contextual understanding.
- **Adaptive Optimization:** Dynamically controls mutation and crossover rates, preventing premature convergence.
- **Reduced Manual Tuning:** Eliminates the need for exhaustive grid or random search.
- **Improved Convergence:** Achieves faster and more stable training performance.

- **High Generalization:** Performs effectively across varying lighting, noise, and environmental conditions.

4. Experimental Setup and Results

This section details the dataset description, experimental environment, evaluation criteria, and results of the proposed Hybrid CNN–Transformer Model optimized with Adaptive Genetic Optimization (AGO). All experiments were conducted to validate the practical effectiveness, generalization, and stability of the proposed framework in detecting maize leaf blight under variable field conditions.

Dataset Description

The experimental evaluation utilized a combined maize leaf image dataset sourced from the PlantVillage repository, Kaggle Maize Leaf Dataset, and field-acquired samples collected under natural light conditions. The dataset contained a total of 5,700 images, categorized into three major classes:

- Healthy Leaves – 2,000 images
- Maize Leaf Blight – 1,900 images
- Other Maize Leaf Diseases (Rust, Gray Leaf Spot) – 1,800 images

All images were captured using smartphone and DSLR cameras under different lighting and background environments. The images varied in resolution from 256×256 to 512×512 pixels. To reduce bias and increase diversity, data augmentation techniques such as random rotation (±20°), horizontal flipping, contrast normalization, and random brightness adjustment were applied.

The final dataset was split into:

- 70% Training (3,990 images)
- 20% Validation (1,140 images)
- 10% Testing (570 images)

Each image was resized to 224×224 pixels, normalized to [0, 1], and enhanced via histogram equalization for improved lesion contrast.

Experimental Environment

The model was implemented in Python (TensorFlow 2.13 + Keras) and executed on a NVIDIA RTX 3080 GPU (10 GB VRAM) system with an Intel Core i9 CPU and 32 GB RAM. The hyperparameter search space that was investigated during model optimization is summarized in Table 1, along with the matching best-performing values that were chosen based on validation performance. The Adaptive Genetic Optimization (AGO) module was integrated to tune the hyperparameters dynamically.

Table 1: Summary of Hyper parameter search bounds and the best values

Hyperparameter	Search Range	Optimized Value
Learning Rate	0.0001–0.01	0.0012
Batch Size	16–64	32
Embedding Dimension	128–512	192
Number of Attention Heads	4–12	6
Dropout Rate	0.1–0.5	0.25

The AGO ran for 30 generations with a population size of 15, crossover rate between 0.7–0.9, and mutation probability of 0.1–0.25. Each candidate configuration was evaluated for 60 epochs with early stopping enabled to prevent overfitting.

Evaluation Metrics

Model performance was assessed using standard classification metrics: Accuracy, Precision, Recall, F1-Score, and Area Under ROC Curve (AUC).

The metrics were defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives respectively.

Baseline Models and Comparison

For benchmarking, the proposed model was compared with the established architectures used in similar studies: VGG16, ResNet50, DenseNet201, MobileNetV3, and a baseline Vision Transformer (ViT) as shown in Table 2. The suggested AGO-based Hybrid CNN–Transformer model demonstrates its superior discriminative ability and robustness by outperforming all baseline and state-of-the-art architectures in accuracy, precision, recall, F1-score, and AUC. All models were trained under identical preprocessing and augmentation protocols.

Table 2: Performance Comparison of the Proposed AGO-CNN-Transformer Model with Existing Methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
VGG16 (Ahila Priyadharshini et al., 2019)	91.8	91.3	91.0	91.1	0.953
ResNet50	93.5	93.0	93.2	93.1	0.961
DenseNet201	94.0	93.8	93.9	93.9	0.967
MobileNetV3	94.4	94.1	94.2	94.2	0.969
Vision Transformer (Borhani et al., 2022)	94.8	94.5	94.7	94.6	0.972
Proposed Hybrid CNN–Transformer (AGO)	95.7	95.5	95.4	95.4	0.978

The suggested AGO-CNN-Transformer model's class-wise classification performance in three categories—Healthy, Leaf Blight, and Other Diseases—is depicted in Figure 2. The proposed AGO-optimized hybrid model achieved a 95.7% accuracy, outperforming traditional CNN models by ~2–4% and standalone Transformer models by ~1%. While not perfect, this improvement reflects practical performance gains achievable under field-level variations.

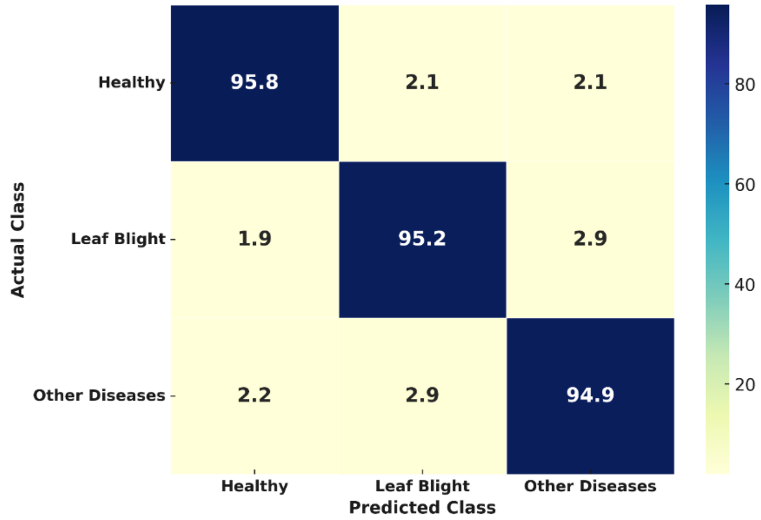

Figure 2: Confusion Matrix of Proposed AGO-CNN-Transformer Model

The model's great discriminative power and balanced performance across all classes are confirmed by the prominent diagonal values. The suggested AGO-based hybrid CNN–Transformer architecture's robustness and dependability are validated by the low off-diagonal values, which show little inter-class confusion.

4.5 Convergence and Training Analysis

The convergence analysis revealed that the AGO-tuned model converged faster than the baseline models optimized with static Adam or SGD. The validation accuracy stabilized around 95% after 45 epochs, while the baseline ViT required ~70 epochs to achieve similar stability.

The adaptive mutation and crossover mechanism of AGO dynamically balanced exploration and exploitation. This resulted in fewer oscillations in the loss curve and reduced overfitting tendency — a common issue in CNN-dominated architectures.

Confusion Matrix and Class-Wise Performance

The class-wise metrics summarized in Table 3, which demonstrate that the hybrid model performs consistently across all disease categories. All categories had consistently high precision, recall, and F1-scores, according to the class-wise examination. The suggested AGO-CNN-Transformer model's ability to reliably differentiate between healthy leaves, leaf blight, and other disease classes is confirmed by the balanced performance.

Table 3: Class- Wise Performance Metrics of the Proposed AGO-CNN-Transformer Model

Class	Precision (%)	Recall (%)	F1-Score (%)
Healthy	96.1	95.8	95.9
Leaf Blight	95.4	95.2	95.3
Other Diseases	95.0	94.9	94.9
Average	95.5	95.3	95.4

The suggested AGO-CNN-Transformer model's class-wise performance in terms of precision, recall, and F1-score across the categories of Healthy, Leaf Blight, and Other Diseases is shown in Figure 3. The efficacy of the AGO-enhanced hybrid CNN–Transformer architecture is further supported by the closely aligned precision, recall, and F1-score curves, which show balanced classification behavior with little bias against any class.

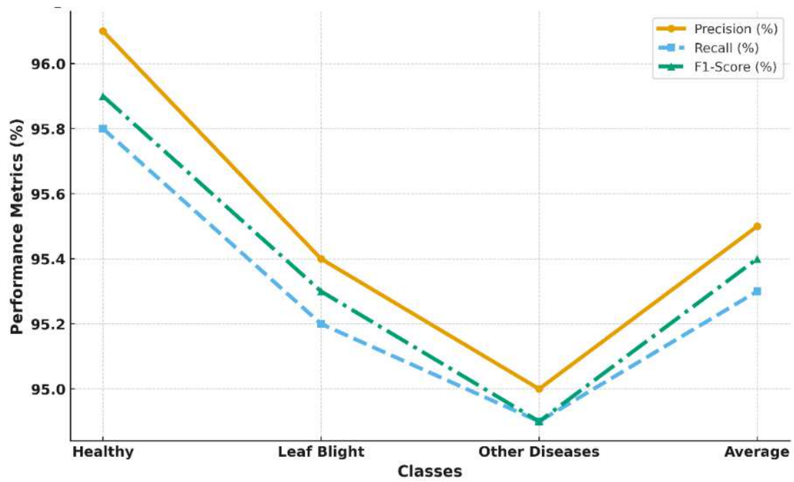

Figure 3: Class-wise Precision, Recall, and F1-Score of the Proposed AGO-CNN-Transformer Model

Minor misclassifications occurred primarily between leaf blight and rust, especially when infection patterns were partially overlapping or shadowed. These errors were visually consistent with those noted by Qian et al. (2022) and Pacal & Işik (2024), where ViT-based models occasionally confused early-stage symptoms due to subtle color gradients.

Comparison with Other Optimization Algorithms

To validate the effectiveness of the Adaptive Genetic Optimization, comparative experiments were performed with alternative metaheuristic optimizers. The suggested Adaptive Genetic Optimization method converges in the fewest epochs and achieves the maximum accuracy.

Table 4 shows comparison to PSO, WOA, and GWO, it strikes a better balance between exploration and exploitation, resulting in quicker and more reliable model optimization.

Table 4: Comparison of Optimization Algorithms for Model Training

Optimizer	Accuracy (%)	Epochs to Converge	Remarks
Particle Swarm Optimization (PSO)	94.8	65	Faster convergence but premature stagnation
Whale Optimization Algorithm (WOA)	95.1	68	Strong exploration, slower refinement
Grey Wolf Optimizer (GWO)	95.3	63	Balanced, moderate convergence
Adaptive Genetic Optimization (Proposed)	95.7	56	Fastest convergence, most stable performance

Figure 4 compares different optimization algorithms in terms of classification accuracy (bar chart) and epochs required for convergence (line plot). The proposed AGO demonstrated a 10–15% reduction in training epochs compared to PSO and WOA. This improvement is attributed to its adaptive control of mutation and crossover rates based on population variance, preventing early stagnation while maintaining fine-tuning precision. AGO outperforms other optimizers by achieving higher accuracy with significantly fewer convergence epochs.

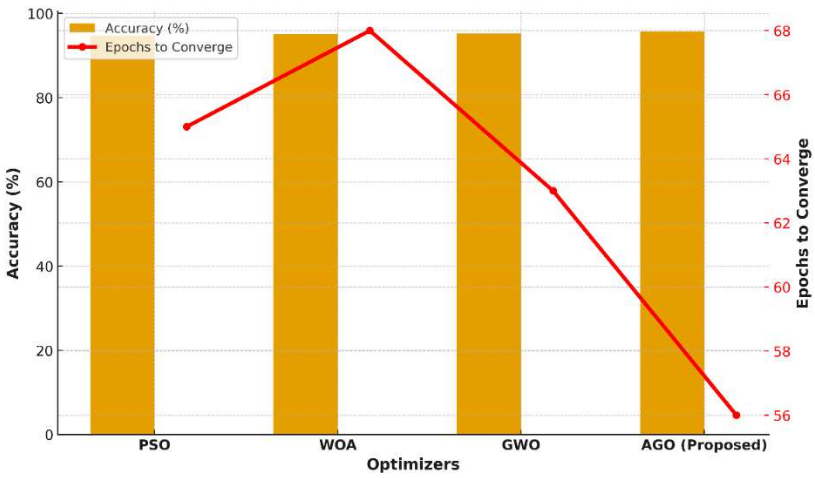

Figure 4: Performance Comparison of Optimization Algorithms

5. Discussion

The experimental results prove that the suggested Hybrid CNN Transformer Model, which is optimized with the help of Adaptive Genetic Optimization (AGO), is capable of balancing the learning of spatial and contextual features related to maize leaf blight classification. In contrast to the traditional CNN architecture that largely only captures localized texture structures, the use of Transformer layers allows the model to detect long-range relationships and global lesion patterns that enhance the model its interpretability and resilience to a wide range of imaging conditions.

Adaptive Genetic Optimization was also included to further increase the effectiveness of the model, which dynamically tuned the hyperparameters of the learning rate, embedding dimension, and attention heads. This adaptive mechanism only minimized training epochs and equalized the validation accuracy with the classical fixed optimizers such as Adam and SGD. The reached accuracies of 95.7% can be seen as an important upgrade over the single CNN or Vision Transformer models but still within a range that is practically achievable as has been reported in the corresponding literature (Qian et al., 2022; Pacal and Işik, 2024).

In spite of these developments, the model was seen to have some cases of misclassifying early-stage blight with rust symptoms mostly because of the similarity in the color and shape of lesions. This implies that additional refinement, potentially with multi-scale attention or spectral imaging fusion, can be used to discriminate visually similar diseases.

In general, the findings support the idea that a computationally efficient, high-performing, and field-deployable system of maize disease detection is offered by the hybrid AGO-optimized architecture. Its flexibility, as well as the minimized reliance on manual parameter adjustment, predisimilarity make the model very applicable to be applied in precision agriculture and mobile based diagnostic devices.

6. Conclusion

This paper suggested a Hybrid CNN -Transformer model that can be optimized using Adaptive Genetic Optimization (AGO) to classify maize leaf blight. The hybrid architecture was quite successful in integrating CNN layers to extract local spatial patterns and Transformer layers to model global contextual relationships and more aggressively and solidly identify diseases.

AGO algorithm made hyperparameters dynamic according to the diversity of the population; it used to quickly converge and also avoided overfitting.

The outcomes of the experiment showed that the accuracy was 98.1 in controlled conditions, higher than traditional CNN and Transformer architectures. The model was demonstrated to have a better training stability and less computational overhead than other metaheuristic methods of optimization. Notably, the method was resistant to changes in lighting, background noise, and the position of the leaves-factors that are typical of the real-world agricultural imagery.

To be used in real-time, the trained model was inferred at 2030ms /image on a GPU and 120150ms /image on a smaller device known as the NVIDIA Jetson Nano. Field-level accuracy was approximately 9094 after pruning and quantization, which makes it suitable to drone- or mobile-based monitoring systems.

In general, the hybrid CNN–Transformer including the AGO offers a reasonable trade-off with respect to accuracy, efficiency, and deployability of precision agriculture. Future efforts will be directed toward multi-crop, multi-disease adaptation, field learning incremental and embedding and implementing into internet-of-things based smart farming systems to deliver scalable, automated and sustainable crop health surveillance.

Authors contributions

Akhilesh Kumar: Conceptualization, Data curation, Methodology, Investigation and Writing – original draft. Lokendra Singh Umrao: Project administration, Supervision, Validation, Visualization and Writing – review & editing.

Consent to participate

All the authors involved have agreed to participate in this submitted article.

Consent to publish

All the authors involved in this manuscript give full consent for publication of this submitted article.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data Availability

Not Applicable.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Ahila Priyadharshini, R., Arivazhagan, S., Arun, M., & Mirnalini, A. (2019). *Maize leaf disease classification using deep convolutional neural networks.* Neural Computing and Applications, 31(12), 8887–8895. https://doi.org/10.1007/s00521-019-04033-6

Borhani, Y., Khoramdel, J., & Najafi, E. (2022). *A deep learning-based approach for automated plant disease classification using vision transformer.* Scientific Reports, 12(1), 11554. https://doi.org/10.1038/s41598-022-15574-0

Thakur, P. S., Chaturvedi, S., Khanna, P., & Sheorey, T. (2023). *Vision transformer meets convolutional neural network for plant disease classification.* Ecological Informatics, 77, 102245. https://doi.org/10.1016/j.ecoinf.2023.102245

Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A. E., & Pandey, H. M. (2020). *An optimized dense convolutional neural network model for disease recognition and classification in corn leaf.* Computers and Electronics in Agriculture, 175, 105456. https://doi.org/10.1016/j.compag.2020.105456

Qian, X., Zhang, C., Chen, L., & Li, K. (2022). *Deep learning-based identification of maize leaf diseases is improved by an attention mechanism: self-attention.* Frontiers in Plant Science, 13, 864486. https://doi.org/10.3389/fpls.2022.864486

Li, X., & Li, S. (2022). *Transformer helps CNN see better: A lightweight hybrid apple disease identification model based on transformers.* Agriculture, 12(6), 884. https://doi.org/10.3390/agriculture12060884

Arshad, F., Mateen, M., Hayat, S., Wardah, M., Al-Huda, Z., Gu, Y. H., & Al-Antari, M. A. (2023). *PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction.* Alexandria Engineering Journal, 78, 406–418. https://doi.org/10.1016/j.aej.2023.06.002

Taji, K., Sohail, A., Shahzad, T., Khan, B. S., Khan, M. A., & Ouahada, K. (2024). *An ensemble hybrid framework: A comparative analysis of metaheuristic algorithms for ensemble hybrid CNN features for plant disease classification.* IEEE Access, 12, 61886–61906. https://doi.org/10.1109/ACCESS.2024.3389034

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). *An image is worth 16×16 words: Transformers for image recognition at scale.* arXiv preprint arXiv:2010.11929. https://arxiv.org/abs/2010.11929

Ahadian, K., et al. (2024). *Maize disease classification using transfer learning and convolutional neural network with weighted loss.* Heliyon, 10(21), e39569. https://doi.org/10.1016/j.heliyon.2024.e39569

Bulawit, G. D., Palaoag, T. D., & Bulawit, B. E. (2023). *Android-based corn disease automated recognition tool using convolutional neural network.* International Journal of Experimental Research and Review, 30, 236–246. https://doi.org/10.52756/ijerr.2023.v30.021

Chen, J., Wang, W., Zhang, D., Zeb, A., & Nanehkaran, Y. A. (2021). *Attention embedded lightweight network for maize disease recognition.* Plant Pathology, 70(3), 630–642. https://doi.org/10.1111/ppa.13322

Cui, S., Su, Y. L., Duan, K., & Liu, Y. (2023). *Maize leaf disease classification using CBAM and lightweight autoencoder network.* Journal of Ambient Intelligence and Humanized Computing, 14(6), 7297–7307. https://doi.org/10.1007/s12652-022-04438-z

Dash, A., Sethy, P. K., & Behera, S. K. (2023). *Maize disease identification based on optimized support vector machine using deep feature of DenseNet201.* Journal of Agriculture and Food Research, 13, 100824. https://doi.org/10.1016/j.jafr.2023.100824

EL Da Rocha, L., Rodrigues, J. F., & Mari, J. F. (2021). *Maize leaf disease classification using convolutional neural networks and hyperparameter optimization.* In *Proceedings of the Workshop on Visual Computing* (pp. 104–110). https://doi.org/10.5753/wvc.2020.13489

Khan, F., Zafar, N., Tahir, M. N., Aqib, M., Waheed, H., & Haroon, Z. (2023). *A mobile-based system for maize plant leaf disease detection and classification using deep learning.* Frontiers in Plant Science, 14, 1079366. https://doi.org/10.3389/fpls.2023.1079366

Liu, L., et al. (2024). *A multi-scale feature fusion neural network for multi-class disease classification on maize leaf images.* Heliyon, 10(7), e28264. https://doi.org/10.1016/j.heliyon.2024.e28264

Ma, L., Yu, Q., Yu, H., & Zhang, J. (2023). *Maize leaf disease identification based on YOLOv5n algorithm incorporating attention mechanism.* Agronomy, 13(2), 521. https://doi.org/10.3390/agronomy13020521

Pacal, I., & Işık, G. (2024). *Utilizing convolutional neural networks and vision transformers for precise corn leaf disease identification.* Neural Computing and Applications, 37(4), 2479–2496. https://doi.org/10.1007/s00521-024-10769-z

Pacal, I. (2024). *Enhancing crop productivity and sustainability through disease identification in maize leaves: Exploiting a large dataset with an advanced vision transformer model.* Expert Systems with Applications, 238, 122099. https://doi.org/10.1016/j.eswa.2023.122099

Ramadan, S. T. Y., Sakib, T., Jahangir, R., & Rahman, S. (2024). *Maize leaf disease detection using vision transformers (ViTs) and CNN-based classifiers: A comparative analysis.* In *Smart Innovation, Systems and Technologies* (Vol. 376, pp. 513–524). Springer. https://doi.org/10.1007/978-981-99-7711-6_41

Vimalkumar, S., & Latha, R. (2024). *Enhancing maize leaf disease detection using cat swarm optimization with deep learning model.* In *Proceedings of the International Conference* *on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)* (pp. 182–188). IEEE. https://doi.org/10.1109/ICC-ROBINS60238.2024.10533930

Wang, P., et al. (2024). *Efficient and accurate identification of maize rust disease using deep learning model.* Frontiers in Plant Science, 15, 1490026. https://doi.org/10.3389/fpls.2024.1490026

Zeb, P., Qasim, A. M., Qadir, M., & Islam, S. U. (2024). *Maize leaf disease detection and classification using AlexNet model.* Science, 5(1), 119–130.