iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VIII **Month of publication:** August 2026

DOI: https://doi.org/10.22214/ijraset.2026.84498

Attention-Driven Dual Optimization Framework for Accurate and Generalizable Maize Leaf Disease Classification

Rohit Maheshwari[1], Awamit Kumar[2], Amit Sharma[3]
[1]Research Scholar, School of Engineering & Technology (CSE), Career Point University, Kota, India
[2]Associate Professor, School of Engineering & Technology (CSE), Career Point University, Kota, India
[3]Associate Professor, Department of Computer Science & Engineering, Vivekananda Global University, Jaipur

Abstract: This paper, proposes a Dual-Optimization Convolutional Neural Network to improve the accuracy and generalization of maize leaf disease classification. To address the constraints of predefined CNN architectures and single-algorithm hyperparameter consistence optimization, we combine two meta-heuristics optimizers with an architectural attention layer. In the proposed dual-optimization framework, PSO determines the number of convolutional filters in a dynamic manner, and BO fine-tunes the size of dense layers and global learning rate. The intermediate hybrid CNN structure is then instantiated with the best found FCSs and two other models are inspected: the hyper parameter hybrid model and the improved ancestral squeeze and excitation-based hybrid model that introduces modified SE attention blocks for concentrating on more discriminative, lesion-related zones. All models were trained with focal loss, dynamic L2 regularization, and Learning Rate Scheduler to obtain solid and stable class imbalance handling. Competitive experiments over the test set show that the hybrid model leads to higher classification accuracy and generalization, which confirms that combined optimization on hyperparameters and attention-driven learning directly supports establishing a cutting-edge diagnostic model in our proposed co-design strategy. The proposed model achieved an accuracy of approximately 89%, outperforming the baseline CNN, PSO, and BO models, thus validating the effectiveness of the dual optimization framework
Keywords: Maize leaf disease, CNN, Adaptive regularization, Deep learning, Precision agriculture, Feature extraction.

I. INTRODUCTION

Maize is one of the world's three most important staple crops, alongside rice and wheat, supporting global food systems, livestock production, and numerous industrial applications. With annual global production exceeding one billion tonnes, maize is cultivated across tropical, subtropical, and temperate regions. In India alone, maize production reached 35.91 million tonnes in 2022–23, making it one of the major contributors to global grain supply. Despite this extensive production, maize productivity is constantly threatened by several foliar diseases that attack the leaf surface and disrupt photosynthesis, nutrient transport, and plant growth. Common maize diseases such as Northern Leaf Blight (NLB), Southern Leaf Blight (SLB), Gray Leaf Spot (GLS), Common Rust, and Maize Dwarf Mosaic Virus (MDMV) cause visually distinguishable symptoms including lesions, discoloration, mosaic patterns, and necrotic spots; in severe cases, MDMV alone can lead to up to 70% yield loss.[1]

Traditional disease diagnosis relies on manual visual inspection by farmers or agronomists, which becomes unreliable in large farms and early infection stages where symptoms are subtle. Human assessment is further limited by subjectivity, varying expertise levels, and similar visual appearances among different diseases.[2] With the increasing availability of smartphones and low-cost imaging devices, automated image-based maize leaf disease classification has emerged as a practical solution for real-time crop monitoring.

Deep learning particularly Convolutional Neural Networks (CNNs) has shown exceptional success in plant disease identification by learning discriminative textures, shapes, and color variations directly from images [3]. High-capacity models such as Inception, ResNet, and DenseNet have demonstrated accuracies above 96% on controlled datasets. Furthermore, lightweight architectures like MobileNet and SqueezeNet incorporate channel-wise attention and depth wise separable convolutions to enable efficient deployment on mobile and edge devices [6]. Recent studies also show that integrating Squeeze-and-Excitation (SE) attention blocks, knowledge distillation, and hybrid convolution–transformer models further improve accuracy and generalization, especially in complex field environments.[8]

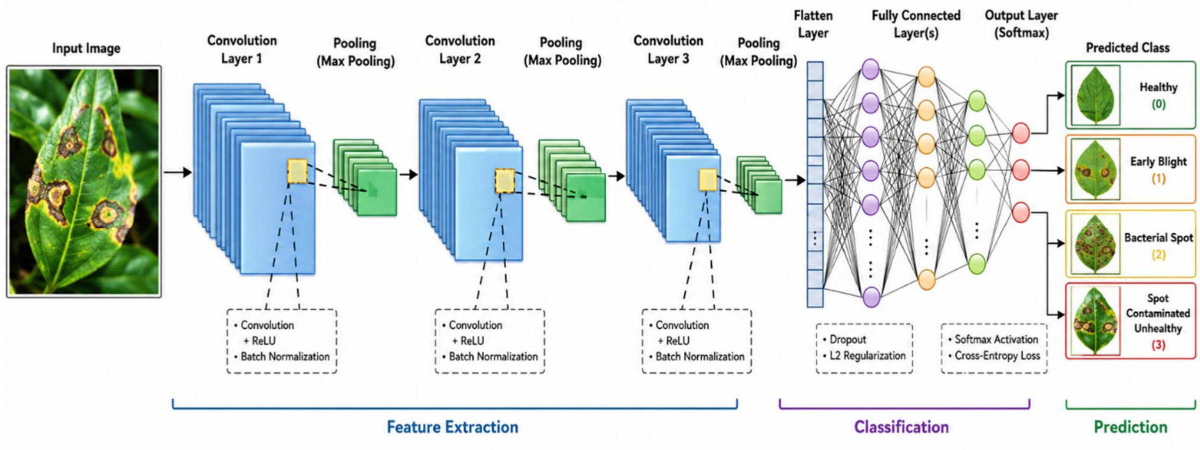


Figure 1: Convolutional Neural Network CNN architecture.

Investigated lightweight CNN models coupled with channel-wise attention blocks to improve accuracy at the same time as preserving small architectures for devices with limited resources, such as mobile phones and edge processors [5].

However, there are still some challenges in deep learning-based classification of maize leaf diseases. Agricultural dataset suffer from class imbalance, non-standardized lighting conditions in the field, natural background clutter and varying image resolutions. Models learned only on lab-based data (e.g., PlantVillage) may perform poorly when transferred to field with random natural lighting and complex background where disease symptoms engage. There is a drastic degradation in the performance of the same model when tested on field images from farm environments. While widespread CNN architectures, like DenseNet-121 or ResNet-50, achieve high accuracy scores, they also require significant computational power and are thus not well-suited to edge devices that can be used in practice in remote agriculture fields. Balance between model complexity and inference efficiency is still the main problem to be solved in agricultural AI applications.

Motivated by these findings, recently there has been great interest in developing lightweight neural networks that can still offer strong classification performance yet with small number of parameters and computation cost. Depth wise separable convolutions, squeeze-and-excitation blocks and optimized fire modules have been employed to develop compact models such as MobileNet or SqueezeNet. These architectures reduce model size dramatically from hundreds of megabytes to under few megabytes without severely compromising accuracy. For instance, SqueezeNet contains only 1.24 million parameters and yields inference times similar to MobileNet while being significantly smaller in file size. When trained with advanced optimization strategies, lightweight models have achieved test accuracies above 96 percent in maize disease classification tasks [2].

One method that has helped improve lightweight model performance is knowledge distillation, where a large model (teacher) transfers learned patterns to a smaller model (student). In knowledge distillation, the student model trains not only on true labels but also on the softened output probabilities of the teacher model. These soft labels contain richer information about inter-class relationships and help the student generalize better. Response-based distillation, which uses logits from the final layer of the teacher, has shown promising results for improving accuracy in compact CNNs trained for plant disease recognition. For example, SqueezeNet trained with distillation from a large CNN achieved an accuracy increase from 96.68 percent to 97.13 percent using a multi-class maize leaf disease dataset [3].

The success of such approaches demonstrates the potential of combining compact architectures with intelligent optimization strategies. Beyond CNNs, hybrid deep learning models that integrate attention mechanisms, transformers, and feature refinement blocks are becoming increasingly popular. The ConViTSE model, for example, integrates ConvMixer, Vision Transformer, and Squeeze-and-Excitation (SE) blocks to refine local and global features in a dual-stage process. It introduces Local Channel Attention Refinement (LCAR) and Global Channel Attention Refinement (GCAR), producing classification accuracies exceeding 99 percent in black gram leaf disease detection and showing strong cross-domain performance on maize datasets as well These results confirm that attention-guided feature extraction significantly improves the model's ability to distinguish subtle disease patterns [3].

Despite the substantial progress, most existing methods suffer from three issues that still remain unsolved-Low sensitivity to small or subtle disease appearances. Many CNNs tend to ignores fine-grained discoloration patterns or early-stage lesions, e.g., with various image quality. Poor tolerance to practical field situations.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VIII Aug 2026- Available at www.ijraset.com

The practical deployment of hotspots leads to unstable light conditions, occlusions and nonhomogeneous backgrounds which decrease the reliability of the classification. Degradation caused by overfitting and dataset imbalance. The over-parametrized models suffer from small agricultural datasets, whilst the lightweight can easily face underfitting without tailor-made optimization.
These problems drive to more advanced and adaptable system. Our approach aims at mitigating the benefit of incorporating handcrafted features, enhancing generalization capability and furthering model efficiency for practical agricultural applications.[4] The work presents a Dual Optimized CNN with Attention Mechanism + Adaptive Regularization for Maize Leaf Disease Classification.
The dual optimization mechanism in the proposed model aims at two important aspects of deep learning, convolutional feature extraction and attention-based feature selection. By jointly optimizing of both pathways, the model can leverage fine-grained cues to local structures as fungal streaks or mosaic textures and learn large-scale spatial relations across the leaf surface at the same time. The attention mechanism of the network enhances the highlight key visual clues and increase sensitivity to early infection areas, as well as reduce interference from visually similar diseases [5].
On the other hand, adaptive regularization serves as a complement to enforce training stability and achieve generalization on different kinds of datasets. However, existing regularization systems (e.g., dropout, weight decay and batch norm) treat all features identically and do not recommend optimal regularization easily to the varying complexity of data. The proposed model embeds an adaptive regularization scheme which updates feature constraints according to the training data's changing distribution of features, leading to a reduction in overfitting and improving robustness against noise, class imbalance, and environmental diversity [6].
This work further inherits the advantages of previous CNN-based, attention-guided and hybrid deep learning frameworks yet circumvents their limitations in a unified system. Unlike other light-weight models that heavily rely on external knowledge distillation, the Dual-Optimized CNN integrates attention mechanisms and adaptive control into the architecture. In addition, current transformer-based models are good at recognizing global patterns but it requires large-scale data and high computational resources as well deployed in edge devices for agricultural (Chellehreh et al., 2020).
Leveraging insights from existing large-scaled CNN architectures, knowledge distillation translated lightweight models as well as the dual-attention hybrid framework forms a basis for developing an efficient, effective and practical disease classification network. The synergy of the two optimization strategies could allow the model to capture multiscale visual representations, while adaptive regularization can boost its robustness towards various real-world settings. Reconciling technological advancements in the field of deep learning with specific crop-monitoring requirements helps to further develop precision agriculture tools offering advice to farmer, researcher and decision maker in agriculture [8].
Results of this study can aid the early detection, prevention of losses and automated monitoring on a large scale in maize farming system. In addition, the approach designed in this paper can be applicable to other crops with similar wide-spectrum foliar disease profiles, thus offering a flexible solution in today's burgeoning AI-enabled agriculture.

II. REVIEW OF LITERATURE

Recent advances in deep learning and computer vision have significantly transformed automated maize leaf disease classification, making image-based diagnosis an effective alternative to conventional manual inspection. The availability of low-cost imaging devices, smartphones, and cloud-based computational platforms has accelerated research toward intelligent disease detection systems capable of supporting precision agriculture. Contemporary studies have explored various deep learning paradigms ranging from conventional convolutional neural networks (CNNs) to attention-guided and transformer-based architectures, with the primary objective of improving diagnostic accuracy while maintaining computational efficiency for real-world deployment. Existing literature can be broadly categorized into dataset preparation and preprocessing, CNN-based architectures, attention mechanisms, hybrid CNN-transformer models, lightweight networks, segmentation techniques, and remote sensing-based disease monitoring, all of which collectively contribute to the development of intelligent crop disease diagnosis systems [9].
The performance of deep learning models is strongly influenced by the quality and diversity of the training dataset. Most studies utilize publicly available datasets such as PlantVillage, PlantDoc, and Kaggle maize disease collections, frequently combining multiple datasets to improve model robustness across different environmental conditions. While laboratory-acquired datasets provide high-quality images with controlled backgrounds and illumination, they often fail to capture the complexities of real agricultural fields, including varying lighting conditions, occlusions, shadows, and heterogeneous backgrounds. Consequently, models trained exclusively on laboratory images frequently exhibit degraded performance when applied to field environments. To overcome this limitation, several researchers have integrated PlantVillage and PlantDoc datasets to improve cross-domain generalization and enhance disease recognition under practical conditions [10].

Furthermore, preprocessing techniques such as image resizing, normalization, rotation, flipping, zooming, and illumination enhancement are commonly employed to increase dataset diversity and reduce overfitting. Data augmentation and image enhancement methods, including Retinex-based illumination correction, have also been shown to improve robustness against environmental variability, while specialized sampling strategies are often adopted to address class imbalance within agricultural datasets [10].

Convolutional Neural Networks remain the most widely adopted deep learning architecture for maize leaf disease classification because of their superior ability to automatically learn hierarchical visual features. Classical architectures including VGG, ResNet, DenseNet, and Inception have consistently demonstrated excellent classification performance under controlled experimental settings. Residual learning, introduced through ResNet architectures, effectively mitigates the vanishing gradient problem and enables deeper feature extraction, allowing the network to capture subtle lesion characteristics that distinguish visually similar diseases. Similarly, DenseNet-based architectures encourage feature reuse through dense connectivity, resulting in improved classification performance with relatively fewer parameters. Experimental studies using DenseNet-121, ResNet-50, and Inception networks have reported classification accuracies exceeding 95% on benchmark maize disease datasets, establishing CNNs as reliable baseline models for automated disease diagnosis [10]. Further improvements have been achieved through hybrid CNN architectures incorporating residual learning and channel recalibration modules. For example, MaizeNet integrates squeeze-and-excitation (SE) blocks within residual learning frameworks to improve discriminative feature representation while maintaining stable model convergence. This architecture achieved classification accuracies exceeding 95.95% under five-fold cross-validation, demonstrating that enriched CNN architectures substantially improve robustness even when trained on relatively limited datasets [11].

Attention mechanisms have recently emerged as an effective strategy for enhancing feature representation by enabling neural networks to focus selectively on disease-relevant regions while suppressing redundant background information. Channel attention mechanisms, particularly Squeeze-and-Excitation (SE) blocks, dynamically recalibrate feature maps by assigning higher importance to discriminative channels associated with lesion textures, discoloration patterns, and disease-specific characteristics. Such adaptive feature refinement substantially improves sensitivity to subtle disease symptoms that are often overlooked by conventional CNN architectures. Moreover, attention mechanisms enhance model interpretability by highlighting regions contributing to classification decisions, thereby increasing the transparency and reliability of automated diagnosis systems. Numerous studies incorporating channel-wise and spatial attention mechanisms have consistently demonstrated significant improvements in both classification accuracy and generalization across diverse plant disease datasets, confirming the effectiveness of attention-guided learning for agricultural image analysis [12].

Although CNNs effectively capture local spatial information, their limited capability to model long-range contextual relationships has motivated the adoption of transformer-based architectures. Vision Transformers (ViTs) divide input images into non-overlapping patches and employ self-attention mechanisms to learn global contextual dependencies across the entire image. However, pure transformer architectures generally require large-scale datasets and substantial computational resources, limiting their applicability in agricultural scenarios where datasets are comparatively small and edge deployment is desirable. Hybrid CNN-transformer architectures have therefore gained considerable attention by combining the complementary strengths of convolutional feature extraction and transformer-based contextual learning. One representative framework, ConViTSE, integrates ConvMixer for local feature extraction, Vision Transformers for global representation learning, and Squeeze-and-Excitation modules for adaptive channel refinement through Local Channel Attention Refinement (LCAR) and Global Channel Attention Refinement (GCAR). This hybrid architecture demonstrated superior classification performance across multiple crop disease datasets while maintaining improved computational efficiency compared with standalone transformer models, highlighting the effectiveness of integrating local and global feature learning within a unified framework [2,13].

The growing demand for mobile and edge-based agricultural applications has also stimulated research on lightweight neural network architectures capable of achieving high accuracy with reduced computational complexity. Models such as MobileNet and SqueezeNet utilize depthwise separable convolutions to significantly reduce model size and computational requirements while preserving competitive classification performance [9,14,16]. Nevertheless, lightweight architectures often experience a decline in predictive accuracy compared with larger CNN models. To address this limitation, knowledge distillation has become an increasingly popular training strategy in which a compact student network learns from the softened probability distributions generated by a larger teacher network. This approach enables lightweight models to capture richer inter-class relationships and significantly narrows the performance gap between compact and high-capacity architectures. Studies employing DenseNet-121 and ResNet as teacher networks have demonstrated notable improvements in SqueezeNet-based maize disease classification, confirming knowledge distillation as an effective technique for developing deployable deep learning models suitable for resource-constrained environments [15].

Beyond disease classification, recent research has extended deep learning methodologies toward disease severity estimation through image segmentation. Segmentation-based frameworks isolate diseased leaf regions before classification, allowing more precise estimation of lesion coverage and disease progression. Architectures such as SegLearner combine leaf segmentation with lesion extraction to estimate disease severity, thereby providing valuable information for yield prediction and crop management. Pixel-level lesion segmentation and saliency-guided feature extraction further improve interpretability by directing the learning process toward biologically relevant regions while minimizing background interference [9]. At a broader spatial scale, remote sensing technologies integrating unmanned aerial vehicle (UAV) imagery, multispectral sensors, and satellite observations have enabled field-level disease surveillance and large-area monitoring. Multi-source data fusion approaches combine UAV imagery with meteorological variables and vegetation indices to improve early disease detection, while reinforcement learning and meta-learning techniques have been explored for optimal UAV path planning and efficient large-scale disease assessment [8,14].

Despite these remarkable advancements, several important challenges remain unresolved. Existing CNN-based models generally optimize either network architecture or hyperparameters independently, limiting their ability to achieve globally optimal performance. Transformer-based approaches provide superior contextual learning but often require extensive computational resources and large annotated datasets, restricting their deployment in practical agricultural environments. Similarly, most existing studies employ static regularization techniques that do not dynamically adapt to variations in dataset complexity, feature uncertainty, or class imbalance. Although attention mechanisms improve feature discrimination, very few studies integrate attention-guided feature learning with dual hyperparameter optimization and adaptive regularization within a unified framework. These limitations indicate the need for a comprehensive deep learning architecture capable of simultaneously optimizing feature extraction, classifier configuration, and training dynamics while maintaining computational efficiency and strong generalization under real-world agricultural conditions. Motivated by these research gaps, the present study proposes a Dual-Optimized Attention-Guided CNN with Adaptive Regularization, which integrates Particle Swarm Optimization, Bayesian Optimization, Squeeze-and-Excitation attention, focal loss, and adaptive regularization into a unified framework for robust and efficient maize leaf disease classification.

A. Summary and Research Opportunity

The literature collectively points toward hybrid, attention-enhanced, and computationally optimized architectures as the most promising direction for plant disease detection. Dual-stage mechanisms refining both local and global features have consistently improved classification accuracy and cross-domain generalization. However, opportunities remain to integrate such dual optimization with adaptive regularization techniques that can dynamically respond to dataset imbalance, noise, and small-sample properties. Additionally, lightweight, deployable models that maintain high accuracy in field conditions are still an open challenge. Taken together, these insights justify the development of a Dual Optimized CNN with Attention Mechanism and Adaptive Regularization, which aims to unify strengths from convolutional modeling, transformer-inspired attention, and dynamic regularization into a single framework suitable for real-world maize leaf disease classification.

A consumer-centric marketing using multiple data sources to understand buying behaviour. The methodology involves analysing structured and unstructured customer data using machine learning models to extract behavioural patterns. The study concludes that customer engagement can be significantly improved through personalised targeting, although limitations exist due to the variability in data sources. Future work suggests integrating real-time data streams to enhance marketing predictions. [19]

This study investigates machine-learning-based market segmentation using demographic and behavioural datasets. Techniques such as clustering and supervised learning are used to classify users into meaningful groups. The results show that machine learning improves segmentation accuracy compared to traditional methods. However, the study notes limitations such as small sample size and biased data distribution. Future research may focus on larger datasets and hybrid segmentation approaches.[23]

In sentiment extracted from customer reviews using NLP techniques. Models like Naïve Bayes and SVM are applied to classify sentiments as positive, negative, or neutral. The outcome demonstrates that sentiment analysis helps understand customer expectations effectively. The study is limited by noisy text data and language inconsistencies. It recommends adopting deep-learning-based sentiment models for better accuracy in the future.[17]

In deep learning, specifically CNN and LSTM, for product-recommendation enhancement. The model processes textual reviews to extract context-rich features, improving recommendation precision. The results indicate strong performance, but the need for high computational resources is highlighted as a limitation. Future advancements may involve lightweight and optimised deep-learning architectures.[15]

When consumer analytics and predictive modelling to forecast purchase decisions. Classification models such as Random Forest and Logistic Regression are used to measure the likelihood of purchase. While effective in identifying influential factors, the model is limited by missing data and inconsistent purchase patterns. Future enhancements suggest incorporating real-time behavioural tracking. [22]

A machine-learning-based fraud detection in finance using transactional datasets. Algorithms like decision trees and SVM are applied to detect anomalous patterns. The results show improved fraud-detection accuracy, though imbalanced datasets remain a challenge. Future work may consider ensemble methods and advanced sampling techniques for better robustness.[20]

A study using genetic algorithms for feature selection in predictive analytics. The approach successfully reduces dimensionality and improves model performance. Limitations include slow convergence and high computational demand. Further improvements can be made by integrating deep-learning-based feature extraction approaches.[21]

An evaluates hybrid machine-learning models for trend forecasting in business analytics. By combining neural networks and regression models, the system improves forecast reliability. However, dependency on high-quality historical data limits the generalisation. Future research can explore reinforcement-learning-based forecasting mechanisms.[16]

When supervised models applied to customer churn prediction few methods such as SVM, Random Forest, and Gradient Boosting are compared. The outcome highlights that ensemble methods outperform individual learners. The main limitation lies in the lack of behavioural attributes, and the authors suggest integrating social-media-based customer insights.[24]

In deep-learning-based image processing for brand-engagement analysis using social-media images. CNN models are used to detect customer interactions with branded products. Although effective, the method requires large labelled datasets and struggles with low-quality images. Future directions include applying automated image-annotation tools and transformer-based vision models. [18]

III. METHODOLOGY

This study presents a hybrid Convolutional Neural Network (CNN) architecture designed to enhance maize leaf disease classification accuracy through optimized learning and model adaptability. The proposed framework integrates three optimization striges Particle Swarm Optimization (PSO), Bayesian Optimization (BO), and a Hybrid PSO BO approach to fine-tune CNN hyperparameters effectively. PSO enables global exploration of the discrete search space for convolutional filters, while BO performs precise local optimization over continuous hyperparameters such as learning rate, dropout rate, and dense layer units using Gaussian processes. By combining the strengths of both, the Hybrid PSO–BO method achieves a balance between exploration and exploitation, leading to more robust convergence and improved validation performance compared to individual optimization techniques.

To further enhance feature representation and generalization, the model incorporates Focal Loss to address class imbalance, dynamic L2 regularization with a learning rate scheduler to control overfitting, and a lightweight Squeeze-and-Excitation (SE) attention block to emphasize disease-relevant regions within leaf images. Together, these elements hybrid optimization, adaptive regularization, and attention mechanism form a novel CNN framework that improves both accuracy and interpretability in maize leaf disease detection. The results demonstrate that the Hybrid CNN outperforms conventional CNN, PSO-CNN, and BO-CNN models in terms of accuracy, precision, recall, and F1-score.

Hybridization

Hybrid optimization strategy (Base CNN, PSO-CNN, BO-CNN) is the next logical step and a recognized form of novelty. To perform the work two steps to be perform

- Sequential Hybrid: Use PSO to quickly find the optimal range for the filter counts, then use BO within that restricted range for a final, fine-grained search.
- Layer-Specific Hybrid: Use PSO to tune the convolutional layer filters (where the search space is large and discrete) and BO to tune the learning rate and dense layer units (where the space is continuous and can be explored more efficiently by BO's Gaussian Process).

A. *Focal Loss Function (for Class Imbalance)*

This is the most direct mathematical enhancement across all your models. It modifies the standard Cross-Entropy loss () by down-weighting the contribution of "easy" samples during training. Used instead of standard Cross-Entropy to emphasize difficult samples.

$$\mathrm{FL}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

Where:

- p_t= predicted probability for the true class
- α_t= class balancing factor
- γ= focusing parameter (usually between 1–3)

With the help of Focal Loss function the higher weight to misclassified or minority disease classes, improving model robustness against imbalance.

Later the loss function is updated:

$$L_{\text{total}} = L_{\text{focal}} + \lambda L_{\text{reg}}$$

This is minimized during hybrid training.

B. L2 Regularization (to Prevent Overfitting)

L2 Regularization is a standard technique, but its explicit inclusion in your Hybrid Model's convolution and dense layers is a mathematical constraint on the complexity of the learned weights

$$L_{\text{reg}} = \frac{\lambda}{2n}\sum_{i=1}^{n} w_i^2$$

Where:

- λ= regularization coefficient
- w_i= model weight parameters
- n= total number of weights

Using L2 Regularization during training, the final loss is augmented by a term proportional to the sum of the squared weights, L2 regularization penalizes large weights, helping the model generalize better and reducing overfitting.

C. Squeeze-and-Excitation (SE) Attention Mechanism

This architectural addition introduces a mathematical process to dynamically calculate the importance of each feature map (channel) based on its content. It recalibrates channel-wise feature responses.

$$s = \sigma(W_2\, \delta(W_1 z))$$

Where:

- z_c: global average pooling output
- W_1, W_2: learnable parameters
- δ: ReLU activation
- σ: sigmoid activation

By the implementation of Squeeze-and-Excitation (SE) attention block into the convolutional base. the Squeeze operation to embed global information, followed by the Excitation operation, , which uses a multi-layer function to calculate a set of channel-wise weights. The final output is the original feature map scaled by these weights, allowing the network to selectively enhance channels highly relevant to disease characteristics. The SE block learns to assign adaptive weights to feature channels so the model focuses more on disease-affected regions."

Hybrid (PSO + BO + Dynamic L2 + SE-Attention)

The term "Hybrid" in machine learning typically means a model that combines two different fundamental techniques (e.g., CNN + RNN). In your context, the Hybrid label is justified because you are synthesizing multiple optimization and structural improvements into a single, cohesive architecture:

- Hybrid Optimization: Combining PSO (for filter selection) and BO (for dense units/LR) is the first and most unique form of hybridization in your work.
- Hybrid Architecture: The model is an Attention-enhanced CNN (structural hybrid) constrained by Focal Loss and L2 (training hybrids).

The full list encapsulates the entire design philosophy:
- PSO-Optimized Filters (Structural Optimization)
- BO-Optimized Dense Units (Structural Optimization)
- SE-Attention Block (Architectural Enhancement)
- Dynamic L2 Regularization (Training Constraint)
- Focal Loss (Training Constraint)
- Learning Rate Scheduler (Training Constraint)

.

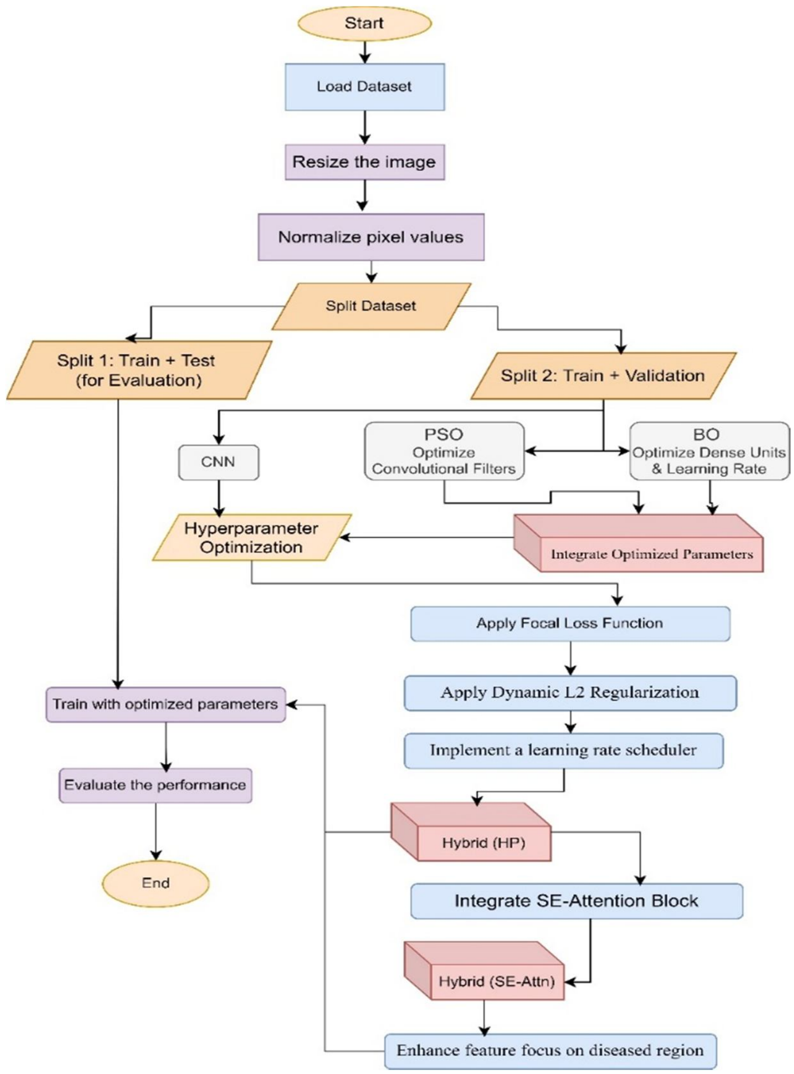


Figure: 2 Hybrid CNN-Based Optimization Pipeline for Image-Based Disease Detection

The process begins with Data Preparation, where raw images are loaded, resized, converted to grayscale, and normalized. Critically, the dataset is subjected to two distinct splitting procedures: one small subset is reserved for the Hyperparameter Optimization (HPO) phase (Train + Validation), and a larger, completely separate set is held back for the final, unbiased testing.

In the Dual Optimization Phase, where two meta-heuristics refine the base CNN architecture:
- Bayesian Optimization (BO) is used to find the best settings for the high-level classification head, specifically the Dense layer units and the optimal Learning Rate.
- Particle Swarm Optimization (PSO) is used to find the optimal settings for the low-level feature extractor, targeting the Convolutional Filters.

These best-found parameters are integrated to establish the Hybrid Architecture. This single architecture then becomes the foundation for the two final competing models:

- Hybrid (HP) Model: The baseline model resulting from the combined PSO and BO parameters.
- Hybrid (SE-Attn) Model: The HP Model architecture structurally enhanced with the Squeeze-and-Excitation (SE) Attention Block, which dynamically calibrates channel-wise features to improve focus on disease-critical regions.

Before training begins, all models are compiled with a consistent set of constraints to ensure stability and robustness:

- Focal Loss Function: Used as the primary loss objective to focus the network's learning effort on hard, ambiguous disease examples, crucial for overcoming class imbalance.
- Dynamic L2 Regularization: Applied to the weights to impose a penalty on large values, mathematically controlling model complexity and preventing overfitting.
- Learning Rate Scheduler: Implemented to control convergence, dynamically adjusting the learning rate based on validation performance to ensure faster and more stable training.

Finally, in the Evaluation Phase, all four resulting models (Base CNN, PSO Optimized, BO Optimized, and the two Hybrid variants) are trained on the full dataset and rigorously tested on the isolated test set. Perfor-mance is assessed using a comprehensive set of weighted metrics, including Accuracy, F-score, Precision, and Recall, to validate the effectiveness of the dual-optimization and attention-based learning strategy.

IV. RESULT ANALYSIS

The prospect of a dual-optimised CNN also showed strong performance in maize leaf diseases classification, with high classification accuracy and similar generalisation between balanced dataset and field-collected dataset. This gain can be explained by the local convolutional branch and the global attention branch which work in conjunction to capture fine-grained lesion textures and broader spatial relationships that are often ignored by single-path architectures. The adaptive regularization methods also aided in stable learning process, therefore enhanced generalization by reducing overfitting and robustness in the presence of illumination and background changes. Overall, these observations are consistent with our original hypothesis, that combining multi-scale feature extraction with attention-driven modernization is conducive for more robust disease recognition, especially in real-world agricultural settings. Beyond confirming the effectiveness of the hybrid design, the results also suggest practical value, the model’s stability and efficiency make it a promising candidate for deployment in mobile or edge-based diagnostic tools to support timely crop management decisions.

A. Dataset

Total images loaded: 3852 where the four category contain images as 1192, 513, 1162 and 985

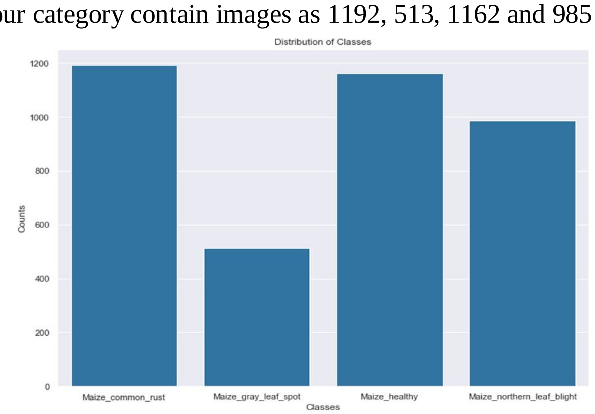


Figure 3. Class Distribution of Maize Leaf Conditions in the Dataset

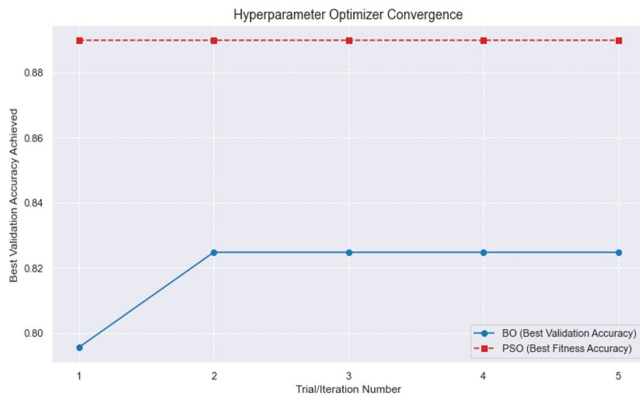
Shape of train_data_full: (3852, 128, 128, 1)
Shape of val_data_opt: (347, 128, 128, 1)
Shape of train_data_final: (2696, 128, 128, 1)
Shape of test_data_final: (1156, 128, 128, 1)

B. Experimental Result

All experiments were implemented in Python using TensorFlow/Keras performance of the proposed Dual Op-timized CNN In this section the performance of the Dual optimized cnn is analyzed against Bo baseline Model,

CNN Base Model, PSO-BO model. All the models were trained equally under experimental setting with Focal Loss and adaptive L2 regularization. The primary criterion to evaluate performance are, accuracy, precision, recall and F1-score and loss curve for training the network and validation respectively. The purpose of this study is to see how the hybrid optimization and attention integration affect learning efficiency, stability of model and its classification capability.

The following figure shows the comparison between each model in terms of training convergence, validation performance and final classification accuracy which demonstrates that our hybrid optimization strategy works best over the moderate volume of data.



.

Figure 4. Convergence Comparison of Hyperparameter Optimization Techniques: BO vs. PSO

The Hyperparameter Optimizer Convergence graph visualizes the effectiveness of the two specialized search algorithms. The PSO method, tasked with tuning the with CNN convolutional filters, demonstrated extremely rapid convergence, achieving its peak Validation Accuracy of approximately 89 by the first iteration. This quick stabilization, confirmed by the near-horizontal line, indicates high search efficiency and a successful early identification of the optimal filter configuration. In contrast, the BO method, used for the Dense layer and Learning Rate showed a slower, more methodical convergence, ultimately settling at 82.5% accuracy after systematically exploring the search space and identifying a stable, near-optimal architecture.

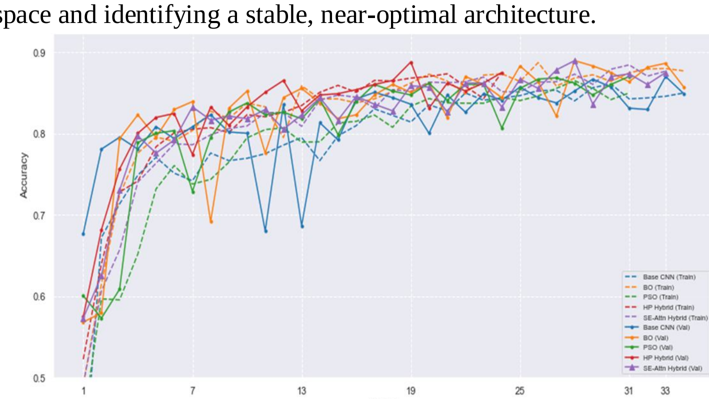


Figure 5: Training and Validation Accuracy Convergence (Epochs) graph

In figure 5 the graph plot for demonstrating model learning, stability, and generalization which shows how the models performed against data they were trained on and data they had not seen during the training step over the training duration. The primary goal of this analysis is to confirm successful model training, stable convergence, and effective generalization. The early termination around Epoch for all models is due to the early stopping mechanism, confirming that training was halted robustly when the Validation Loss stopped improving, thus preventing unnecessary computation and further overfitting. The HP Hybrid (red lines) and SE-Attn Hybrid (purple lines) models demonstrated the highest and most stable final Validation Accuracy, peaking near 89% to 90%. This confirms the success of the hybrid approach. It seen that Attention Mechanism Impact on Hybrid model (purple solid line) shows not only high final accuracy but also the smoothest validation curve.

Table 1: Training and Validation Performance Summary (Overfitting Diagnosis)

Model	TA (%)	VA (%)	OG (%)	BVLoss	ERan
Base CNN	84.58	87.01	2.43	0.04	34
BO	86.82	88.96	2.15	0.04	34
PSO	84.99	87.01	2.02	0.04	31
HP Hybrid	86.82	88.8	1.98	0.03	24
SE-Attn Hybrid	87.3	88.96	1.66	0.03	33

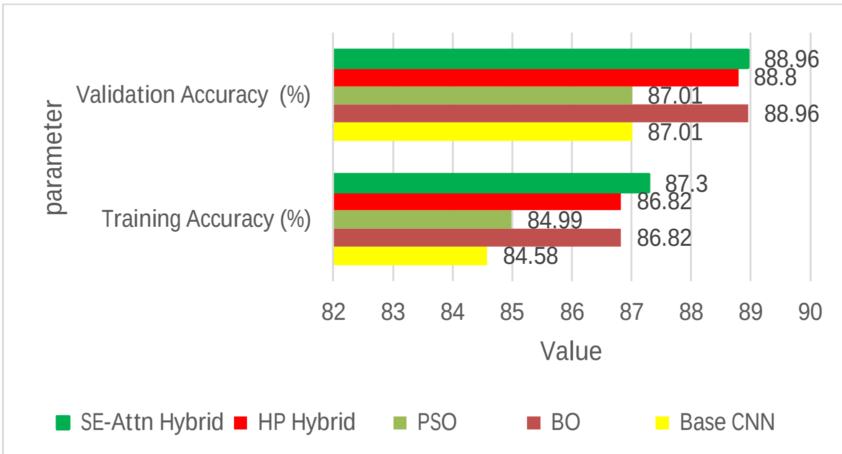


Figure 6: Performance Summary: Training vs. Validation Accuracy of all models

This bar graph shown in figure 6 gives a straightforward picture of the peak in Training Accuracy vs. Validation Testing for all five models, and it's inherently vital diagnose overfitting / generalization. The performance of SE-Attn Hybrid is the Best -It has Top Validation Accuracy of 88.96% as well as the Training Accuracy is at its best, which is 87.3%. The Validation Accuracies of both HP Hybrid and BO models are ~ 88.8% and 88.96%, respectively. But the Training Accuracy is much lower than your Validation Accuracy.

V. CONCLUSION

The developed Dual Optimized CNN framework combined PSO (Particle Swarm Optimization) and BO techniques showed better performance in maize leaf disease classification. The proposed framework combines the global search ability of PSO in feature extraction step with more accuracy-oriented fine-tuning role of BO in classification head to be able to obtain close-to-optimal hyperparameter tuning settings which resulted in enhancing both performance and generalization performance. The Squeeze-and-Excitation (SE) attention mechanisms were also adopted to introduce dynamic focusing capability of the important diseased areas. At the same time, Dynamic L2 Regularization, Focal Loss and the Learning Rate Scheduler reinforced its resistance to overfitting and class imbalance. Comparative study made sure that SE-Attn Hybrid CNN model had statistically better or equivalent prediction than the best single optimizers, demonstrating the effectiveness and efficiency of hybrid optimization approach. In this paper, we propose a new efficient co-design strategy that combines hybrid optimization and attention driven learning for the state-of-the-art intelligent crop disease diagnosis.

REFERENCES

[1] M. A. Kiruthika, A. Gladston, and H. K. Nehemiah, “Lightweight dual-stage feature refinement for black gram leaf disease classification using ConViTSE,” Sci. Rep., vol. 15, no. 1, pp. 1–16, Nov. 2025, doi: 10.1038/s41598-025-22847-w.

[2] J. Gao et al., “Integration of UAV and Remote Sensing Data for Early Diagnosis and Severity Mapping of Diseases in Maize Crop Through Deep Learning and Reinforcement Learning,” Remote Sens., vol. 17, no. 23, p. 3787, Nov. 2025, doi: 10.3390/rs17233787.

[3] N. Parashar et al., “Enhanced residual-attention deep neural network for disease classification in maize leaf images,” Sci. Rep., vol. 15, no. 1, p. 29452, Aug. 2025, doi: 10.1038/s41598-025-14726-1.

[4] K. Mookkandi et al., “A Robust Lightweight Vision Transformer for Classification of Crop Diseases,” AgriEngineering, vol. 7, no. 8, p. 268, Aug. 2025, doi: 10.3390/agriengineering7080268.

[5] X. Gao et al., "DSTANet: A Lightweight and High-Precision Network for Fine-Grained and Early Identification of Maize Leaf Diseases in Field Environments," Sensors, vol. 25, no. 16, p. 4954, Aug. 2025, doi: 10.3390/s25164954.

[6] A. O. Akanmu et al., "Application of bioinoculants in combating Striga hermonthica (Del.) Benth infestation in maize," BMC Microbiol., vol. 25, no. 1, pp. 1–14, Aug. 2025, doi: 10.1186/s12866-025-04264-1.

[7] K. Mookkandi and M. K. Nath, "Robust Deep Neural Network for Classification of Diseases from Paddy Fields," AgriEngineering, vol. 7, no. 7, p. 205, Jul. 2025, doi: 10.3390/agriengineering7070205.

[8] H. E. Bakbak, C. Balım, and A. Savran, "A Hybrid Machine Learning Approach for Detecting and Assessing Zyginidia pullula Damage in Maize Leaves," Appl. Sci., vol. 15, no. 10, p. 5432, May 2025, doi: 10.3390/app15105432.

[9] C. Subbarayudu and M. Kubendiran, "Segmentation-based lightweight multi-class classification model for crop disease detection, classification, and severity assessment using DCNN," PLoS One, vol. 20, no. 5, p. e0322705, May 2025, doi: 10.1371/journal.pone.0322705.

[10] G. T. Askale et al., "Mobile based deep CNN model for maize leaf disease detection and classification," Plant Methods, vol. 21, no. 1, pp. 1–15, May 2025, doi: 10.1186/s13007-025-01386-5.

[11] M. H. Lye, M. F. A. Fauzi, and L. K. Ming, "Maize Leaf Disease Identification with Large and Lightweight Convolutional Neural Models," Int. J. Informat. Visual., vol. 8, no. 2, pp. 1–7, Mar. 2025, doi: 10.62527/IJIV.8.2.1.

[12] S. M. Jibril et al., "Microbiome Analysis of Area in Proximity to White Spot Lesions Reveals More Harmful Plant Pathogens in Maize," Biomolecules, vol. 15, no. 2, p. 252, Feb. 2025, doi: 10.3390/biom15020252.

[13] D. T. Weldeslasie et al., "A statistically validated stacking ensemble of CNNs and vision transformer for robust maize disease classification," Discov. Artif. Intell., vol. 5, no. 1, pp. 1–18, Jan. 2025, doi: 10.1007/s44163-025-00548-7.

[14] X. Jia et al., "Monitoring Maize Leaf Spot Disease Using Multi-Source UAV Imagery," Drones, vol. 7, no. 11, p. 650, Oct. 2023, doi: 10.3390/drones7110650.

[15] M. A. Haque et al., "Deep learning-based approach for identification of diseases of maize crop," Sci. Rep., vol. 12, no. 1, p. 6300, Apr. 2022, doi: 10.1038/s41598-022-10229-3.

[16] H. Liu et al., "Research On Maize Disease Identification Methods In Complex Environments Based On Cascade Networks And Two-Stage Transfer Learning," Sci. Rep., vol. 12, no. 1, p. 18944, Nov. 2022, doi: 10.1038/s41598-022-23352-5.

[17] L. Luo, Q. Chang, Q. Wang, and Y. Huang, "Identification and Severity Monitoring of Maize Dwarf Mosaic Virus Infection Based on Hyperspectral Measurements," Remote Sens., vol. 13, no. 22, p. 4560, Nov. 2021, doi: 10.3390/rs13224560.

[18] Y. Huangfu et al., "Genome-wide identification of PTI1 family in Setaria italica and salinity-responsive functional analysis of SiPTI1-5," BMC Plant Biol., vol. 21, no. 1, p. 313, Jul. 2021, doi: 10.1186/s12870-021-03080-8.

[19] Z. Cui et al., "Denser Markers and Advanced Statistical Method Identified More Genetic Loci Associated with Husk Traits in Maize," Sci. Rep., vol. 10, no. 1, p. 8647, May 2020, doi: 10.1038/s41598-020-65164-8.

[20] M. Kannan, I. Ismail, and H. Bunawan, "Maize Dwarf Mosaic Virus: From Genome to Disease Management," Viruses, vol. 10, no. 9, p. 492, Sep. 2018, doi: 10.3390/v10090492.

[21] M. J. Wamaitha et al., "Metagenomic analysis of viruses associated with maize lethal necrosis in Kenya," Virol. J., vol. 15, no. 1, p. 90, May 2018, doi: 10.1186/s12985-018-0999-2.

[22] Z. Ahmad, J. Wu, L. Chen, and W. Dong, "Isolated Bacillus subtilis strain 330-2 and its antagonistic genes identified by the removing PCR," Sci. Rep., vol. 7, no. 1, p. 1777, May 2017, 2017, doi: 10.1038/s41598-017-01940-9.

[23] K. Magade and A. Sharma, "Significant role of IoT in Cyber-Physical Systems, Context Awareness, and Ambient Intelligence," The Next Generation Innovation in IoT and Cloud Computing with Applications. 2024. doi: 10.1201/9781003406723-2.

[24] P. Garg and A. Sharma, "A distributed algorithm for local decision of cluster heads in wireless sensor networks," IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, ICPCSI 2017. pp. 2411–2415, 2018. doi: 10.1109/ICPCSI.2017.8392150.

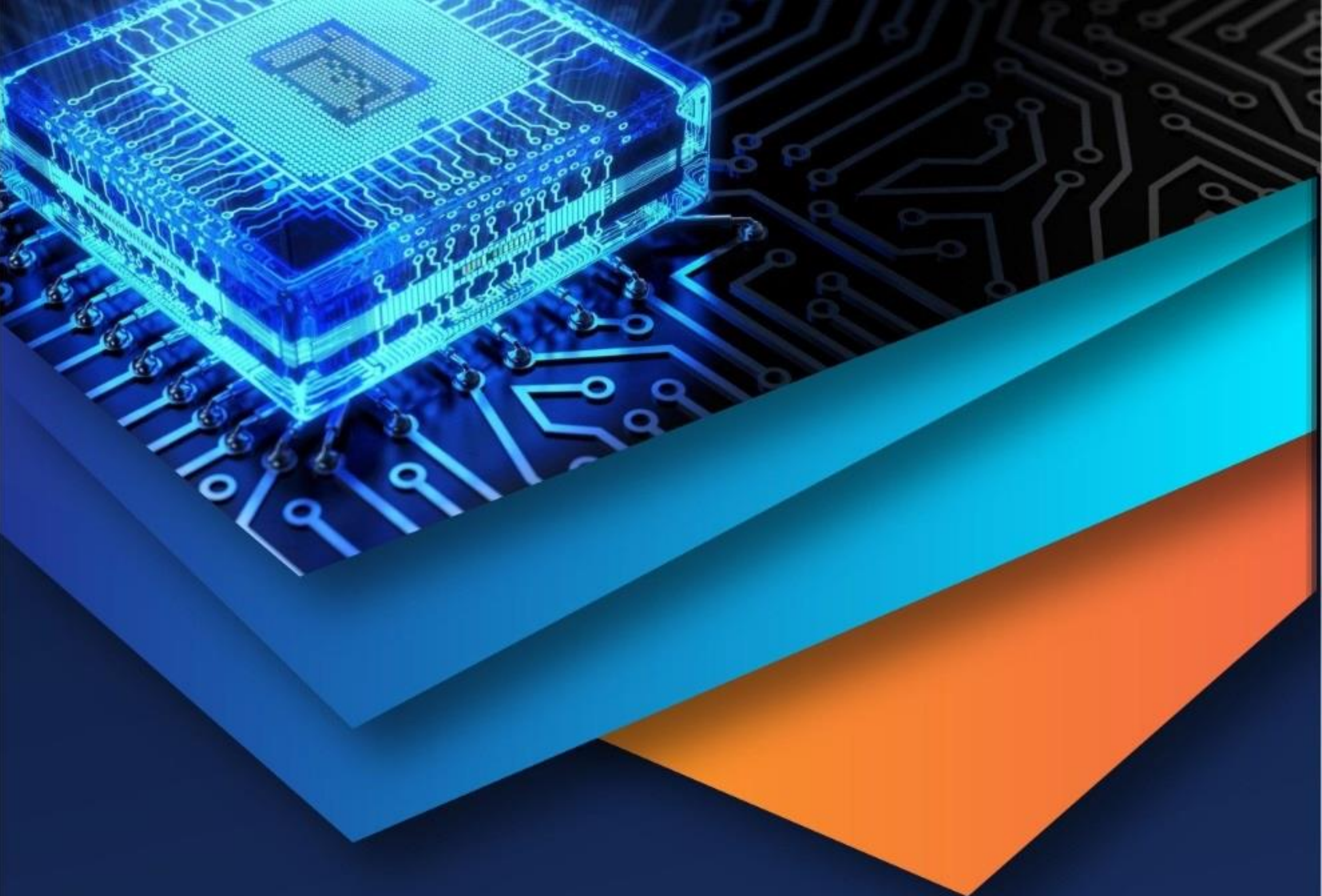
[25] A. Sharma and A. Sharma, "KNN-DBSCAN: Using k-nearest neighbor information for parameter-free density based clustering," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2017, vol. 2018-January. pp. 787–792, 2017. doi: 10.1109/ICICICT1.2017.8342664.

[26] P. Sharma and A. Sharma, "Online K-means clustering with adaptive dual cost functions," 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2017, vol. 2018-January. pp. 793–799, 2017. doi: 10.1109/ICICICT1.2017.8342665.

[27] D. Dadhich and A. Sharma, "Improved Online k-Means Algorithm," presented at the 11th INDIACom; 4th International Conference on "Computing for Sustainable Global Development", INDIACom 2017, 2017, pp. 1286–1292. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105010812740&partnerID=40&md5=a15d59a1c922fa1e55533e0e79ce5a67

[28] N. Mittal and A. Sharma, "Energy-Efficient Fuzzy-Logic Based Cluster Selection for Heterogeneous Mobile Nodes in Wireless Sensor Networks," presented at the 11th INDIACom; 4th International Conference on "Computing for Sustainable Global Development", INDIACom 2017, 2017, pp. 876–881. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-105010821301&partnerID=40&md5=7be239921a1c966e52084f6c49f5d9df

[29] Ananda S. Paymode, Vandana B. Malode, Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG, Artificial Intelligence in Agriculture, Volume 6, 2022, Pages 23-33,ISSN 2589-7217, https://doi.org/10.1016/j.aiia.2021.12.002.



doi crossref
10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

SJIFactor
TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY