Banana Leaves Nutrient Deficiency Detection using Latent Attention Convolutional Neural Network

V. Rekha[1], Uma Shankari Srinivasan[2]

Abstract

Early and accurate diagnosis of nutrient deficiencies in banana leaves is crucial for ensuring optimal crop health and maximizing yield. We present the Convolutional Latent Attention Network (CLAN), a unified deep learning framework that synergistically combines global context modeling and region-based localization for precise classification of nutrient disorders. CLAN begins with a lightweight hierarchical convolutional encoder that extracts multi-scale feature maps from high-resolution leaf images, culminating in a compact latent code. By applying self attention, the Latent Attention Refinement module accentuates key global features linked to deficiencies. The Region Proposal Network (RPN) operates on the highest-level encoder features to identify candidate areas of interest. These regions are pooled via ROIAlign across all encoder stages and processed through a MobileNetV2 backbone to generate detailed local descriptors. By concatenating the refined latent code with region-wise features, the CLAN classification head achieves robust identification of multiple deficiency classes. The proposed model achieves a 95.2% F1 score and processes images at 4.5 ms inference time, while remaining compact at only 1.97 million parameters. These results demonstrate that CLAN combines rapid convergence and strong generalization with real-time capability and minimal hardware demands, making it exceptionally suitable for deployment in field conditions and resource-constrained environments.

[1]Research Scholar, Department of Computer Science and Applications, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, 600089, Tamil Nadu, India

[2]Associate Professor, Department of Computer Science and Applications, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, 600089, Tamil Nadu, India

Email: [1]rekhaonmail@gmail.com [2]umabalajees@gmail.com

Keywords - Deep learning, Banana nutrient deficiency, Latent Attention Network, Region Proposal Network

1. Introduction

Banana (Musa spp.) is the cornerstone of agricultural economies and food systems in tropical and subtropical regions around the world, with a particularly profound impact in countries like India. Its importance extends beyond the mere production of agricultural products, including critical roles in ensuring food security, promoting nutritional stability, and promoting economic viability for millions of small farmers and agricultural communities [1]. In India, banana cultivation holds a prominent position, ranking first in production and third in area among fruit crops, contributing substantially to the nation's agricultural GDP [2, 3]. The crop serves as a vital source of carbohydrates, vitamins and minerals, making it an indispensable component of the dietary intake of a significant portion of the population, especially in rural areas where it often acts as a staple food [4]. Beyond its nutritional value, banana cultivation provides substantial income opportunities and employment, supports livelihoods and contributes to regional economic development [5]. Therefore, sustained growth and productivity of banana crops are paramount for the well-being of these communities and the overall agricultural landscape. Despite its immense importance, banana cultivation is frequently hampered by various challenges, among which nutritional deficiencies are seen as a critical factor that limits optimal yields and crop quality [6].

Banana plants, which are fast growing and high-yielding, have a substantial de- mand for macro and micronutrients throughout their growth cycle. When these essential nutrients are not adequately supplied or absorbed, plants exhibit characteristic symptoms of deficiency, often manifesting as identifiable patterns on their leaves [7]. These visual cues, such as discoloration, stunted growth, or necrotic lesions, are direct indicators of the underlying physio- logical stress caused by nutrient imbalances [8]. If left unaddressed, these deficiencies can severely affect plant health, leading to reduced photosynthetic efficiency, impaired fruit development, and ultimately a significant decline in both the quality and quantity of harvest [9]. The economic repercussions for farmers can be substantial, underscoring the urgent need for a timely and accurate diagnosis of these nutritional disorders.

Historically, the diagnosis of nutritional deficiencies in banana plants has mainly been based on traditional methods, primarily visual inspection by experienced agronomists or farmers. This approach involves a careful examination of leaf symptoms, comparing them with known patterns associated with specific nutrient deficiencies [10, 11]. In addition, effective visual diagnosis requires a high level of expert knowledge and experience, which is often scarce in many agricultural regions. These inherent drawbacks highlight the need for more objective, efficient, and reliable diagnostic tools. In recent years, the agricultural sector has witnessed a transformative shift toward automated solutions, driven by advances in artificial intelligence and machine learning. Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a powerful paradigm for image-based detection of plant disease and deficiency [12]. CNNs have the remarkable ability to learn complex patterns directly from raw image data, which makes them well suited to analyze visual symptoms on plant leaves. This has led to the development of numerous automated systems that aim to overcome the limitations of traditional diagnostic methods. However, despite substantial progress and promising results, existing CNN-based approaches often face their own set of challenges. These include limitations in achieving consistently high accuracy, particularly in diverse field conditions, and a lack of interpretability that makes it difficult to understand the basis of their diagnostic decisions [13].

A key contributing factor to these limitations is often the inadequate integration of global contextual information with precise localization capabilities, which are crucial for a comprehensive understanding of nutrient deficiency patterns. Addressing the aforementioned limitations, this paper proposes a novel framework Convolutional Latent Attention Network (CLAN). CLAN is specifically designed to improve the diagnostic accuracy and applicability of automated nutrient deficiency detection in banana plants by effectively combining global contextual information with precise regional analysis. The CLAN lies in its ability to leverage attention mechanisms to focus on salient features indicative of deficiencies while simultaneously incorporating broader contextual cues from the entire leaf image.

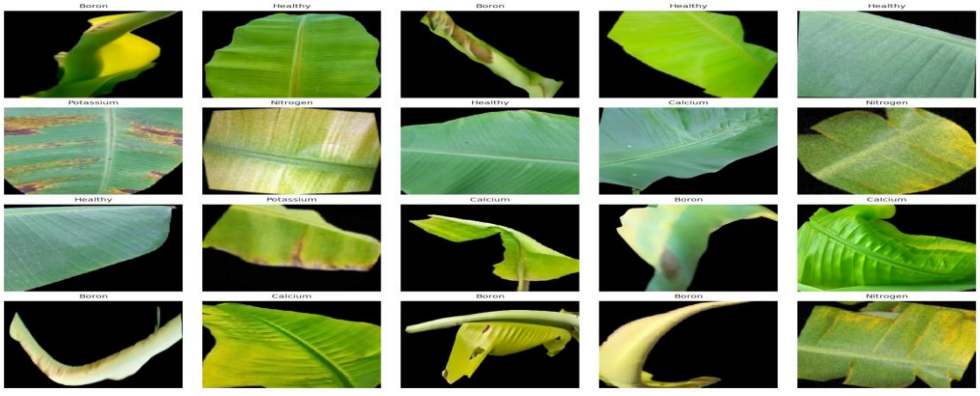

Figure 1. Banana Leaf: Symptoms of healthy leaf and observed Deficiency, including (a) Boron Deficiency, (b) Calcium Deficiency, (c) Potassium Deficiency, and (d) Nitrogen Deficiency

This dual approach allows CLAN not only to pinpoint the exact location of symptoms but also to interpret them within the general physiological state of the plant, leading to a more robust and reliable diagnosis. By integrating these capabilities, the aim of CLAN is to provide a more accurate, interpretable, and practical solution for early and precise identification of nutrient deficiencies, which significantly contributes to improving the quality and quantity of banana yields and supporting sustainable agricultural practices. The sample images of banana leaf deficiency are shown in Figure1.

2. Related Works

Deep learning methods have been increasingly applied to the diagnosis of nutrient deficiencies across a wide range of crops. Watchareeruetai et al. [14] introduced a novel CNN methodology that partitions black gram leaf images into spatial blocks, training specialized CNNs for each to identify five types of macronutrient deficiencies. Their model also enables severity estimation, demonstrating potential for precise, non-invasive plant diagnostics.

Ghosal et al. [15] leveraged a deep convolutional network on RGB images of soybean leaves and proposed an Integrated Framework for Identification, Classification, Quantification, and Prediction (ICQP). This framework facilitates high-throughput phenotyping and stress hotspot mapping, underlining CNNs' capacity for mechanistic detection of abiotic stress. Employing transfer learning, Wulandhari et al. [16] fine-tuned an Inception-ResNet model pretrained on ImageNet on crop images, achieving 96% training accuracy and 86% testing accuracy. Their find- ings underscore the importance of transfer learning and hyperparameter optimization (e.g., learning rate, number of epochs) for field-level agricultural monitoring.

Lewis and Espineli, [17] addressed eight dis- tinct nutrient deficiencies in coffee plants via conventional preprocessing, segmentation, and a customized CNN classifier. High in situ accuracy suggests that CNN pipelines are scalable across varied cropping systems. In a hydroponic rice study, Xu et al. [18] ap- plied DenseNet-121, achieving a validation accuracy of 98.62±0.57% and test accuracy of 97.44±0.57%. The model outperformed ResNet, Inception-v3, and NasNet-Large, confirming deep CNNs' robustness in symptom recognition.

Anami et al. [19] trained a VGG-16 model on a 30,000-image dataset compris- ing five rice cultivars, classifying 12 biotic and abi- otic stress categories. Their CNN surpassed conventional backpropagation networks, demonstrating scal- ability in diverse stress detection tasks. Sethy et al. [20] compared several architectures such as AlexNet, VGG16/19, GoogleNet, ResNet18/50 each paired with a support vector machine (SVM). They found that ResNet-50 coupled with SVM achieved the highest accuracy for nitrogen-deficiency detection and recommended model expansion and updated leaf-color charts to enhance scalability.

Sharma et al. [21] introduced an ensemble transfer learning approach for rice nutrient detection, utilizing public datasets from Mendeley and Kaggle. They combined Inception-ResNetV2 (achieving 90% accuracy) with Xception (95.83% accuracy), illustrating that multi-architecture ensembles can significantly improve nutrient stress prediction. Jayasiri et al. [22] implemented two deep learning paradigms for tomato nutrient deficiency detection: Mask R-CNN for pixel-wise segmentation and YOLO for rapid bounding-box classification. Their systems achieved 92% and 98% accuracy, respectively, and provided spatial distribution estimates to support actionable agronomic decisions.

In a recent study, Sathyan and Palanisamy, [23] em- ployed CNNs augmented with extensive data augmentation on the real-world IPNI leaf dataset. They combined classification with a stochastic gradient descent- based localization strategy, producing high accuracy and efficient convergence, highlighting the effectiveness of augmentation and optimization-driven training for enhanced field robustness. Mkhatshwa et al. [24] conducted a comparative evaluation of CNN backbones - including a standard CNN, VGG 16 and InceptionV3 in rice and banana data sets. Although InceptionV3 achieved the highest precision (93%), the novelty of the study lies in the incorporation of explainability techniques (SHAP and GradCAM) into the evaluation pipeline, effectively balancing performance with interpretability.

Supreetha et al. [25] further demon- strated the power of transfer learning and hybrid architectures by combining pretrained CNNs (InceptionV3, VGG16/19, ResNet50/152) with SVM classifiers. Their ensemble achieved up to 99.05% accuracy in identifying nitrogen, calcium, and potassium (NPK) deficiencies in rice, underscoring the significant potential of such approaches for high-performance agronomic image analysis.

Deep learning has revolutionized agricultural image processing by automating disease and nutrient deficiency detection tasks. CNN architectures, including VGGNet, ResNet, and EfficientNet, have achieved notable success due to their hierarchical feature extraction capabilities. However, these architectures often overlook the critical global context and lack explicit mechanisms for adaptive feature selection [26]. Attention mechanisms have emerged as promising enhancements to CNN architectures. Approaches such as Squeeze and Excitation Networks (SE-Net) and the Convolutional Block Attention Module (CBAM) have successfully demonstrated improved feature recalibration capabilities by dynamically emphasizing informative features [27].

However, their limited scope of application at intermediate layers restricts their capacity to enhance global features. Region-based convolutional neural networks (R-CNN, Fast R-CNN, Faster R-CNN) have shown superior performance in object localiza- tion tasks, primarily through candidate region proposals [28, 29]. These approaches typically operate independently of comprehensive global context analysis, which is crucial for accurate classification of complex nutrient deficiency patterns. CLAN integrates the strengths of hierarchical CNNs, latent global con- text modeling, and region-specific localization, address- ing existing limitations through an innovative combined architecture that significantly enhances both interpretability and classification accuracy.

3. Methodology

In this study, a comprehensive dataset of banana leaf images was utilized to train a neural network to classify nutrient deficiencies in banana plants. The dataset consists of 600 images showing various deficiencies in nutrients in banana leaves, including healthy samples and deficiencies in nitrogen, calcium, potassium, and boron. The images were collected from the Theni and Thanjavur districts of Tamil Nadu, together with additional samples obtained from a public dataset available in Mendeley [30]. The proposed CLAN architecture integrates hierarchical convolutional feature extraction, attention-based global context encoding, and region-specific localization to improve the precision of nutrient deficiency and disease classification in agricultural images as shown in Figure 2. The methodology is organized into several key stages.

3.1. Input Preprocessing

3.1.1. Image standardization

The initial step involves standardizing all input images to a resolution of 256×256 pixels with three color channels (RGB).

Equation (1) for image standardization, each input image $\mathrm{I} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{C}}$ is resized to a fixed resolution

$$\mathrm{I}' = \mathrm{Resize}(\mathrm{I}, 256 \times 256 \times 3) \quad (1)$$

Where H, W are the original height and width, C refers RGB channels, and I' is the standardized image.

3.1.2. Data Augmentation Techniques

To enhance the model's robustness and generalization capabilities, standard data augmentation techniques are applied. These include random horizontal and vertical flips, slight rotations (±15°), brightness and contrast adjustments, and scaling transformations. This preprocessing ensures that the model is exposed to a diverse set of image variations, reducing overfitting and improving its performance on unseen data.

Equation (2) and (3) are used for random horizontal and vertical flips.

For horizontal flip,

$$\mathrm{I}''(\mathrm{x},\mathrm{y},\mathrm{c}) = \mathrm{I}'(\mathrm{W}-\mathrm{x}-1,\mathrm{y},\mathrm{c}) \quad (2)$$

For vertical flip,

$$\mathrm{I}''(\mathrm{x},\mathrm{y},\mathrm{c}) = \mathrm{I}'(\mathrm{x}.\mathrm{H}-\mathrm{y}-1,\mathrm{c}) \quad (3)$$

Where x, y are pixel coordinates and c is the channel index.

Equation(4) is used for random rotation (±15°). Rotation by angle $\theta\epsilon[-15°, +15°]$ is done by

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (4)$$

Here, interpolation is applied to compute pixel intensities at non-integer coordinates.

Equation (5) is used to adjust the brightness of an image and the brightness factor $\beta \epsilon\ [0.8, 1.2]$.

$$I''(x,y,c) = \beta \cdot I'(x,y,c) \quad (5)$$

Equation (6) is used to adjust the contrast of image and the contrast factor $\alpha\epsilon\ [0.8, 1.2]$.

$$I''(x,y,c) = \alpha \cdot (I'(x,y,c) - \mu_c) + \mu_c \quad (6)$$

Where μ_c is the mean intensity of channel c.

Equation (7) for scaling transformation, and the scaling factor $s\ \epsilon\ [0.9, 1.1]$.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \quad (7)$$

The final augmented dataset is,

$$D_{aug} = \{A(I') | I \in D\}$$

where $A(I)$ is the augmentation pipeline applying random flips, rotations, brightness/contrast changes, and scaling. This preprocessing pipeline ensures that the model sees diverse variations of each image, improving generalization and reducing overfitting.

3.2. Light weight Hierarchical Convolutional Encoder

The encoder architecture is designed to progressively extract multiscale features from the input images. It comprises four sequential convolutional blocks, each consisting of two 3×3 convolutional layers followed by batch normalization and ReLU activation functions. After each block, a 2×2 max-pooling layer is applied to reduce the spatial dimensions by half, while doubling the number of filters: 32, 64, 128, and 256, respectively. This hierarchical structure enables the network to capture both fine-grained details and high-level semantic information, producing feature maps F_1 - F_4, with decreasing spatial resolutions and increasing channel depths.

Given an input image, $I \in R^{256\times256\times3}$, the operations within each block can be expressed as,

$$Z = I * K + b \quad (8)$$

Equation (8) as convolution and Equation (9) as batch normalization.

$$\hat{Z} = \frac{Z-\mu}{\sigma}\gamma + \beta \quad (9)$$

To apply ReLU activation Equation (10) as,

$$A = max(0, \hat{Z}) \quad (10)$$

For pooling the equation (11) be as,

$$\mathrm{F_k} = \mathrm{MaxPool(A)} \quad (11)$$

Where K and b denote convolutional kernels and biases, μ, σ are the batch statistics, and γ, β are learnable BN parameters.

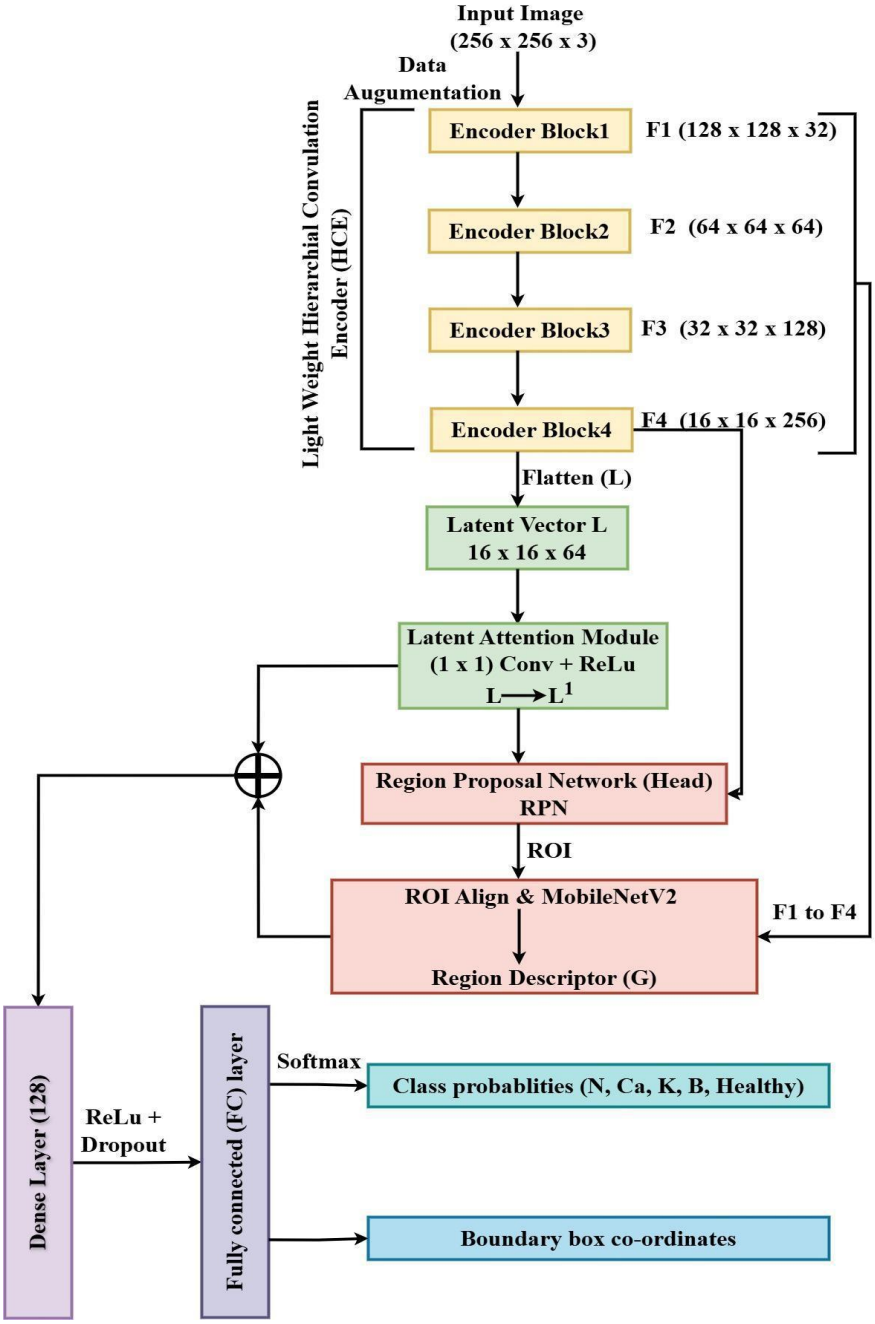

Figure 2. Architecture of CLAN

Through this hierarchical progression, the encoder produces feature maps $F_1 \in R^{128\times128\times32}$, $F_2 \in R^{64\times64\times64}$, $F_3 \in R^{32\times32\times128}$, $F_4 \in R^{16\times16\times256}$

This structure ensures that low-level fine details are captured in early layers F_1, while high-level semantic abstractions emerge in deeper layers F_4. The resulting multi-scale feature hierarchy is crucial for downstream tasks such as classification or segmentation, as it balances spatial precision with semantic richness.

3.3. Bottleneck and Latent Projection

At the deepest level of the encoder, the feature map F_4 with dimensions 16×16×256 is passed through a 1×1 convolutional layer, reducing the channel depth to 64 and generating a compact latent tensor L and the equation (12) as,

$$L = F_4 * K_{1\times1}^{64} + b \qquad (12)$$

Where $K_{1\times1}^{64}$ denotes the set of 1×1 kernels producing 64 output channels, and b represents the bias term.

The resulting latent representation is $L \in R^{16\times16\times64}$. This latent representation serves as a global semantic code that encapsulates the most prominent features relevant to nutrient deficiencies. The dimensionality reduction via the 1×1 convolutional layer ensures computational efficiency while retaining essential information for subsequent processing. L serves as a compact yet information-rich descriptor, particularly effective for capturing discriminative patterns relevant to nutrient deficiency detection.

3.4 Latent Attention Refinement Module

The Latent Attention Refinement (LAR) module is integral to the CLAN architecture, enhancing its capacity to capture and refine global contextual information essential for accurate detection of nutrient deficiency in banana leaves. Initially, the model processes the output of the final encoder layer, producing a feature map of dimensions 16×16×256 that encapsulates high-level semantic features extracted from the input image, as illustrated in Figure 3. To distill this rich feature map into a more compact representation, a 1×1 convolutional layer is applied, reducing the channel depth from 256 to 64. This operation results in a latent tensor that serves as a global semantic code, capturing the essence of the input image's content. Subsequently, three separate 1×1 convolutional layers are used to generate the query (Q), key (K), and value (V) matrices of the latent tensor. Each of these projections produces a tensor of dimensions 16×16×64, allowing the model to learn distinct representations of the latent space. Then these Q, K, and V tensors are flattened spatially, transforming them into matrices of size 256×64, facilitating the computation of attention scores at all spatial locations.

The attention mechanism computes the relationships between each spatial position by calculating the scaled dot product of the query and the key matrices. Mathematically, the attention output is given by:

The mathematical equation (13) is as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (13)$$

where d_k is the dimensionality of the key vectors (64 in this case). The scaling factor $\sqrt{d_k}$ is introduced to prevent the dot products from growing too large, which could lead to extremely small gradients and impede learning. The resulting attention output, a matrix of size 256×64, is then reshaped back to the original spatial dimensions of 16×16×128. This reshaped tensor represents the refined latent features with a greater emphasis on the most salient spatial information. To integrate this refined representation with the original latent tensor, a residual connection is used. The original latent tensor is added element-wise to the attention output, and the combined tensor is passed through a 1×1 convolutional layer followed by a ReLU activation function.

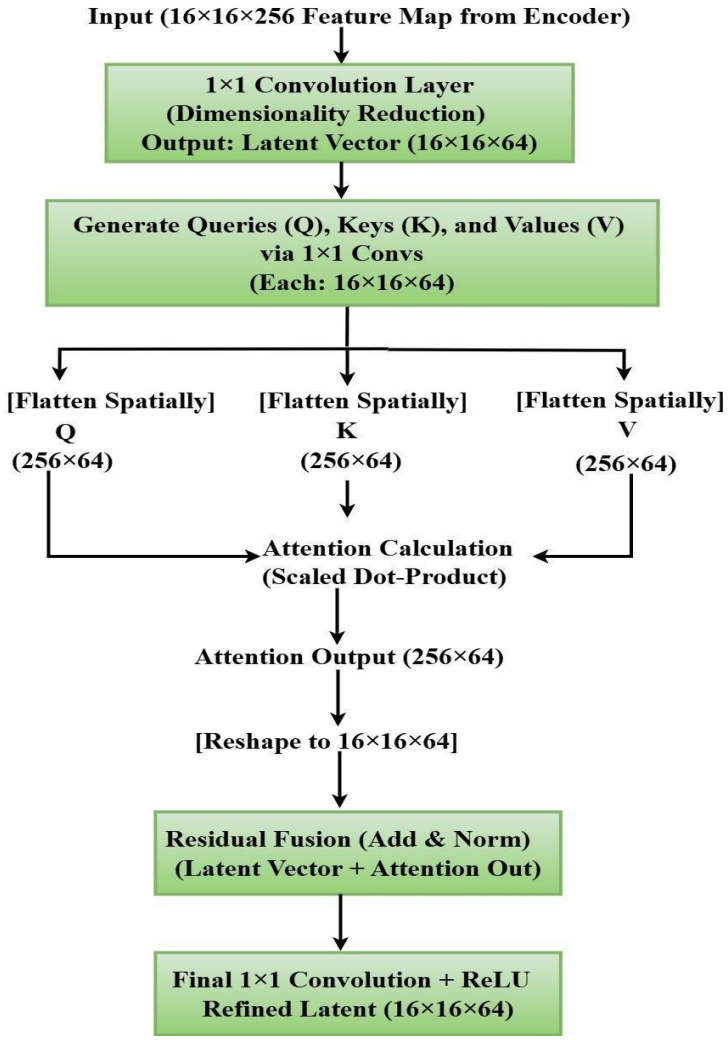

Figure 3. Latent Feature and Attention Module Block Diagram

This residual fusion mechanism facilitates the flow of gradients during training and helps to preserve the original characteristic information. The output of the residual fusion step is a refined la- tent tensor of dimensions 16×16×64. This refined representation captures the most discriminative global features, which are crucial for the

subsequent classification and localization tasks in the CLAN architecture. By focusing on the most informative aspects of the latent space, the LAR module enhances the model's ability to accurately detect and classify nutrient deficiencies in banana leaves.

3.5 ROIAlign and MobileNetV2 Descriptor

In the CLAN architecture, the RPN generates candidate regions of interest (RoI) that potentially contain pertinent information on nutrient deficiencies in banana leaves. To ensure that these regions are represented uniformly, irrespective of their original sizes or aspect ratios, the RoIs are subjected to a Region of Interest Align (ROIAlign) operation. ROIAlign addresses the misalignment issues inherent in traditional RoI pooling by eliminating quantization errors during the feature extraction process. It achieves this by performing bilinear interpolation to compute the exact values of the input features at four regularly sampled locations in each sub-window of the RoI, thereby preserving spatial precision. This results in a fixed-size feature map for each RoI, typically of dimensions 7×7 or 14×14, depending on the subsequent network requirements.

Following ROIAlign operation, the pooled feature maps are passed through a pre-trained MobileNetV2 (width multiplier $\alpha = 0.35$), excluding its fully connected layers. This backbone serves as a robust feature extractor, leveraging its deep residual learning frame- work to capture high-level semantic information pertinent to each RoI. The mobileNetV2 backbone output is then subjected to a global average pooling operation. This pooling technique computes the average of all spatial locations in the feature map, resulting in a fixed-length feature vector that encapsulates the most salient information from the RoI. This fixed length descriptor is later utilized in the classification and localization heads of the CLAN architecture, facilitating the accurate identification and delineation of various classes of nutrient deficiency. By integrating ROIAlign with MobileNetV2, the CLAN model effectively captures detailed local characteristics while maintaining spatial integrity, thus improving its diagnostic capabilities in precision agriculture applications.

3.6. Feature Fusion and Prediction Heads

The refined latent tensor L1 and the region descriptors are concatenated to form a unified feature vector. This vector is passed through a fully connected layer with 128 units, followed by ReLU activation and dropout for regularization. The network then branches into two parallel prediction heads: one for classifying the type of nutrient deficiency (e.g., nitrogen, calcium, potassium, boron, or healthy) using a softmax activation function and another for bounding box regression to refine the coordinates of the detected regions. This dual output structure enables the model to perform classification and localization tasks simultaneously.

4. Results and Discussion

The CLAN architecture introduces an efficient and compact approach to detect nutrient deficiencies in banana leaves. CLAN integrates a lightweight hierarchical encoder, a latent self-attention module for global feature refinement, and a region-based processing branch using MobileNetV2 as the backbone. The CLAN framework was executed on a workstation that features two NVIDIA GeForce RTX2080Ti GPUs, each equipped with 11GB of GDDR6 memory and 64GB of RAM to facilitate efficient processing of banana leaf images. The data set was divided into 80% training sets and 20% validation sets. Trained from end to end on a single 1.97 million parameter footprint, CLAN was rigorously evaluated against four established pre-trained models, VGG16, InceptionV3, DenseNet121 and ResNet50, using a banana nutrient deficiency dataset. All models were trained for 100 epochs under identical conditions: Adam optimizer (β_1=0.9, β_2=0.999), learning rate of 0.001, and batch size of 8.

4.1 Training and Validation Performance

During the training process, the CLAN exhibited rapid convergence. By epoch 100, its training accuracy reached 96.5%, while the validation accuracy at 95.1%, indicative of a strong generalization as shown in Table 2. VGG16 showed early overfitting, achieving 94.0% training accuracy but only 90.0% on validation. InceptionV3 achieved 95.0% training and 92.0% validation accuracy, showing smoother convergence and reduced overfit compared to VGG16. DenseNet121 showed maximum training and validation accuracies of 93.0% and 88.8%, respectively. ResNet50 performed similarly to InceptionV3 in training (95.0%) but lagged in generalization (90.6%). The validation loss curves reinforce the robustness of CLAN. Its loss decreased sharply from 0.80 to 0.15, maintaining a small gap (< 0.05) with training loss. However, other models with higher validation losses (0.18 - 0.30) and a noticeable divergence from the training trajectories, pointing to a weaker generalization.

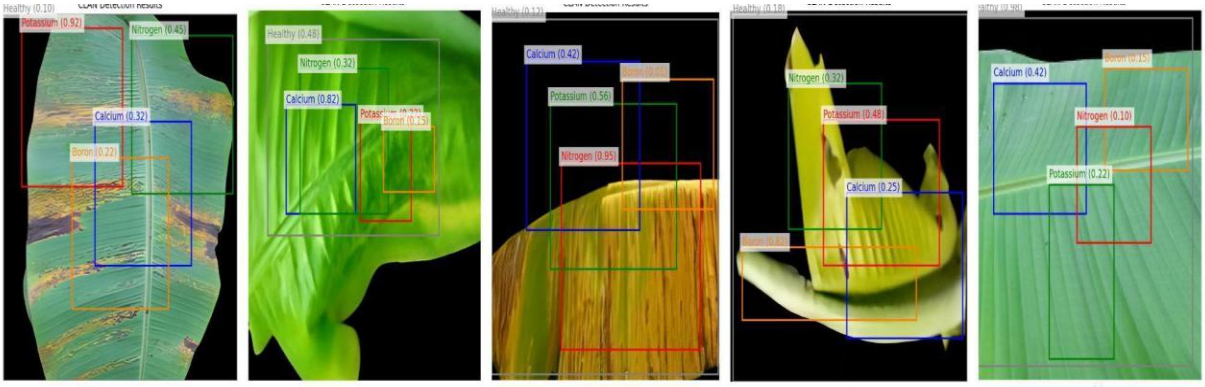

Figure 4. CLAN nutrient-deficiency detection results

The final outputs of the CLAN architecture include, Class probabilities: A vector indicating the likelihood of each nutrient deficiency class for each proposed region. Bounding-box coordinates: Refined coordinates specifying the location of the detected deficiency regions within the input image. These results facilitate the precise identification and localization of nutrient deficiencies, providing actionable information for agricultural applications as shown in Figure 4.

4.2 Performance Analysis Comparison

Table 1: Comparison of Precision, Recall, F_1-Score and Inference Time

Model	Precision (%)	Recall (%)	F-Score (%)	Inference Time (ms/img)
CLAN (proposed)	95.4	95.1	95.2	4.5
VGG16	90.2	89.8	90.0	6.8
InceptionV3	92.4	92.0	92.2	7.5
DenseNet121	89.1	88.6	88.8	8.1
ResNet50	90.8	90.4	90.6	7.2

For a comprehensive evaluation of performance, we computed precision, recall, and F_1 scores across the five nutrient deficiency classes, in addition to measuring the average inference time using a Tesla T4 GPU, as shown in Table 1. CLAN achieved a precision of 95.4%, a recall of 95.1%, and a F_1 score of 95.2%, outperforming all other models. InceptionV3 followed with a 92.2% F1 score, while VGG16 and ResNet50 showed around 90%. DenseNet121 achieved the lowest F_1 score at 88.8%. In terms of efficiency, CLAN processed each image in approximately 4.5ms, outperforming the pre-trained mod- els: VGG16 (6.8ms), ResNet50 (7.2ms), InceptionV3 (7.5ms), and DenseNet121 (8.1ms).

4.3 Confusion Matrix Insights

Figure 5 presents the confusion matrices for the five class nutrient deficiency classification task. The VGG16 model demonstrated strong performance in the detection of multiclass nutrient deficiency. It correctly identified 108 of 120 nitrogen-deficient and 105 of 120 calcium-deficient leaves. However, there were notable mislabeling patterns: 6 nitrogen-deficient leaves were classified as calcium and 6 as potassium, while calcium-deficient leaves were labeled as nitrogen (7) and potassium (8). The potassium-deficient leaves also showed ambiguity, with 4 classified as nitrogen and 10 as calcium. In contrast, boron-deficient and healthy leaves were classified with high precision, showing only two mislabels for boron and a single mislabel for healthy. In the InceptionV3 model, nitrogen-deficient leaves were correctly identified 110 out of 120, 5 misclassified as calcium, and 5 as potassium. Calcium- deficient leaves were correctly classified 108 out of 120, but showed 5 misclassifications as nitrogen and 7 as potassium.

Boron-deficient leaves were nearly perfectly recognized, with just one misclassification in the healthy category, while all healthy leaves were correctly classified. DenseNet121, however, showed the highest misclassification among nitrogen deficiencies: Only 105 of 120 nitrogen-deficient leaves and 103 of 120 calcium-deficient leaves were correctly identified, and potassium was misclassified more frequently. However, both boron-deficient and healthy leaves remained perfectly recognized. ResNet50 produced moderate performance, 107 of 120 nitrogen-deficient leaves and 106 of 120 calcium-deficient leaves were correctly identified. Potassium-deficient samples had 13 misclassifications (4 labeled nitrogen and 9 as calcium), while boron- deficient leaves, two mislabeled as healthy. while healthy classification remained flawless. In contrast, the CLAN model achieved maximum accuracy, reducing errors among closely related classes: only five total misclassifications for nitrogen and calcium, seven for potassium, and perfect precision for boron and healthy categories. These results show that CLAN significantly enhances the distinction between nutrient deficiencies, particularly for the classes of nitrogen, calcium, and potassium that are more visually similar, underscoring its robust and balanced diagnostic performance.

Confusion matrix analysis reveals that CLAN markedly reduces misclassification among the nutrient-deficient

classes (Nitrogen, calcium, potassium), whereas the pre-trained models (especially DenseNet121) exhibited frequent misclassification in these categories. In particular, all models showed the highest precision in identifying boron deficiency and healthy leaves, likely due to their distinct visual symptoms. The CLAN model demonstrates a strong balance between high accuracy and computational efficiency. It delivers 95% accuracy with under 1.97 million parameters, embodied in a practical trade-off suitable for mobile or embedded agricultural applications. It achieves the highest F1 score among all evaluated architectures (95.2%), underscoring its effective handling of class imbalances. Achieving 4.5ms per image makes CLAN an ideal candidate for real-time in-field nutrient monitoring. Operating with only 1.97M parameters, CLAN is exceptionally lightweight and efficient.

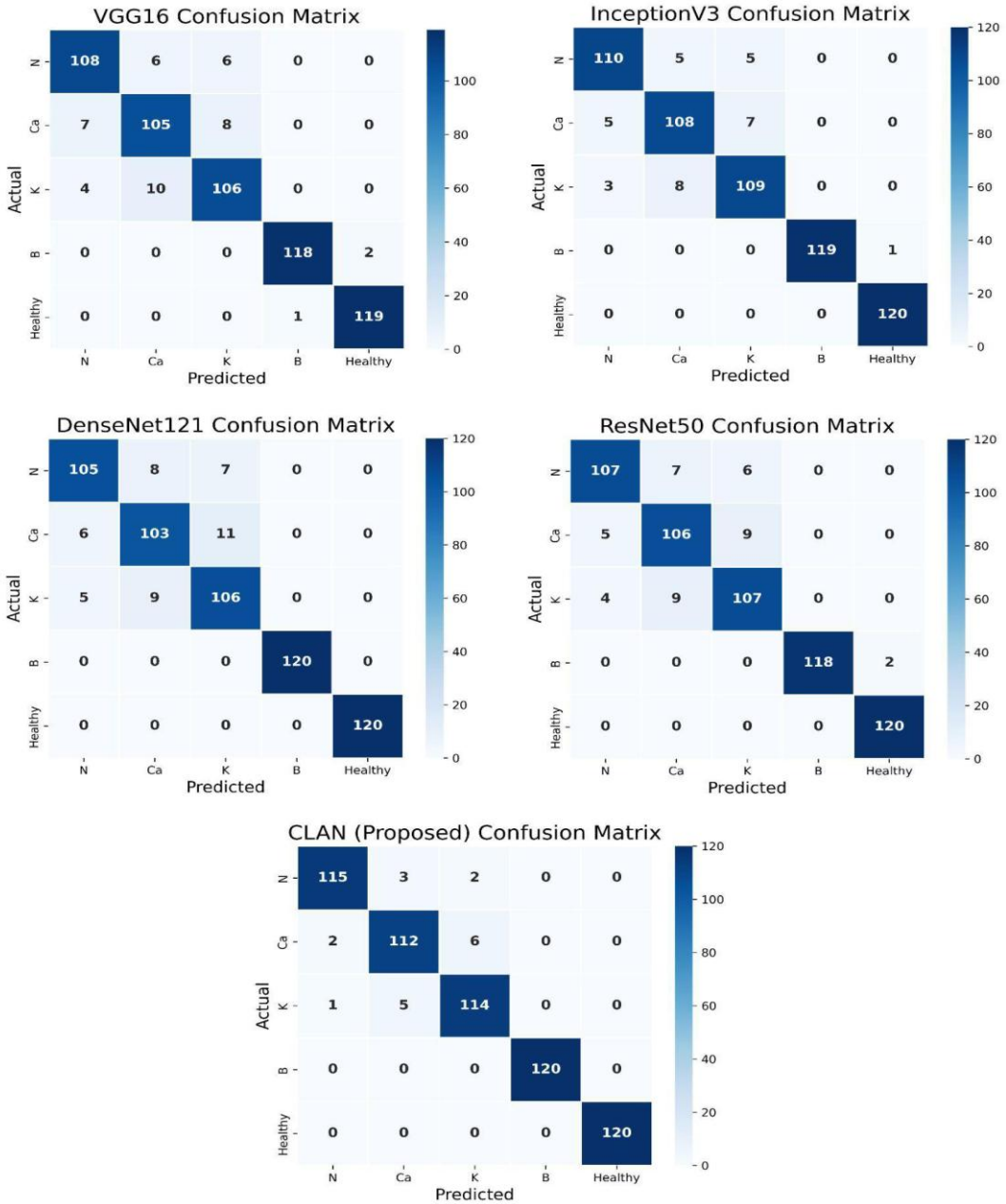

Figure 5. Confusion matrices of nutrient deficiency classification using different CNN models.

Each matrix shows the number of correctly and incorrectly classified samples for five classes.
In general, CLAN successfully combines self-attentive latent feature refinement with region-based processing to achieve state-of-the-art performance in the detection of nutrient deficiency, while remaining suitable for deployment in resource-constrained environments.

4.4 Model Complexity and Parameters Efficiency

The CLAN model consists of approximately 1.97 million parameters, distributed in the encoder, latent attention module, RPN, MobileNetV2 backbone (α= 0.35), and fusion heads, as shown in Table 2. This lightweight configuration supports real-time inference while maintaining strong detection accuracy, making the model suitable for deployment on devices with limited computational resources.

Table 2: Parameter count breakdown for each component in the CLAN architecture

Component	**Parameters (approx.)**
Encoder	~1.56 M
Latent & Attention	~20 K
RPN (Region Proposal Network)	~170 K
MobileNetV2	~190 K
Fusion & Heads	~33 K
Total	≈**1.97 M**

6. Conclusion

In this study, we demonstrate that CLAN significantly outperforms pre-trained backbones in both convergence speed and generalization across unseen leaf samples. These gains stem from its combination of latent attention, which efficiently isolates key symptom regions, compact encoder layers, and region proposal mechanism for precise localization, inspired by Faster RCNN. Lightweight and yet powerful, CLAN processes full resolution leaf images with significantly lower computational overhead, making it suited for deployment on edge devices such as smart phones or UAVs. This aligns with the growing emphasis of precision agriculture on lightweight and accurate detectors in field conditions. Furthermore, the CLAN region-based attention and grouping strategy improves detection accuracy, surpassing the performance of single-stage architectures in agricultural scenarios. These findings underscore the strong potential of CLAN as a practical, field-deployable solution that delivers nutrient deficiency detection that is fast, reliable and resource efficient.

References

1. F. Mrope and N. Jeeva, “Modeling the transmission dynamics of banana bunch top disease in banana plants,” Eurasian Journal of Mathematics, Computer Applications, vol. 12, pp. 73-90, 2024.
2. U. Das and R. K. Bhattacharyya, “Fruit quality of Grand Naine (AAA) banana as influenced by varied components of precision farming system,” International Journal of Chemical Studies, vol. 7, no. 3, pp. 4475-4478, 2019.
3. S. Uma and P. S. Kumar, “Banana research and development in India-challenges and opportunities,” Progressive Horticulture, vol. 52, no. 1, pp. 1-11, 2020.
4. M. M. A. N. Ranjha, S. Irfan, M. Nadeem, and S. Mahmood, “A comprehensive review on nutritional value, medicinal uses, and processing of banana,” Food Reviews International, vol. 38, no. 2, pp. 199-225, 2022.
5. H. Wardhan, S. Das, and A. Gulati, “Banana and mango value chains,” in Agricultural Value Chains in India: Ensuring Competitiveness, Inclusiveness, Sustainability, Scalability, and Improved Finance, pp. 99-143, 2022.
6. S. Muthusamy and S. P. Ramu, “IncepV3Dense: Deep ensemble based average learning strategy for identification of micro-nutrient deficiency in banana crop,” IEEE Access, vol. 12, pp. 73779-73792, 2024.
7. J. D. Thiagarajan, S. V. Kulkarni, S. A. Jadhav, A. A. Waghe, S. P. Raja, S. Rajagopal, H. Poddar, and S. Subramaniam, “Analysis of banana plant health using machine learning techniques,” Scientific Reports, vol. 14, no. 1, p. 15041, 2024.
8. K. Nyombi, “Diagnosis and management of nutrient constraints in bananas (Musa spp.),” in Fruit Crops, Elsevier, pp. 651–659, 2020.
9. A. Naorem, S. Jayaraman, Y. P. Dang, R. C. Dalal, N. K. Sinha, C. S. Rao, and A. K. Patra, “Soil constraints in an arid environment-challenges, prospects, and implications,” Agronomy, vol. 13, no. 1, p. 220, 2023.
10. M. G. Selvaraj, A. Vergara, H. Ruiz, N. Safari, S. Elayabalan, W. Ocimati, and G. Blomme, “AI-powered banana diseases and pest detection,” Plant Methods, vol. 15, pp. 1-11, 2019.
11. P. Sunitha, B. Uma, A. G. Kiran, S. Channakeshava, and C. S. S. Babu, “A convolution neural network with skip connections (CNNSC) approach for detecting micronutrients boron and iron deficiency in banana leaves,” Journal of Umm Al-Qura University for Engineering and Architecture, pp. 1-19, 2024.
12. M. Shoaib, B. Shah, S. Ei-Sappagh, A. Ali, A. Ullah, F. Alenezi, T. Gechev, T. Hussain, and F. Ali, “An advanced deep learning models-based plant disease detection: A review of recent research,” Frontiers in Plant Science, vol. 14, p. 1158933, 2023.
13. A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, “A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope,” Sustainability, vol. 15, no. 7, p. 5930, 2023.
14. U. Watchareeruetai, P. Noinongyao, C. Wattanapaiboonsuk, P. Khantiviriya, and S. Duangsrisai, “Identification of plant nutrient deficiencies using convolutional neural networks,” in Proc. Int. Electrical Engineering Congress (iEECON), pp. 1-4, IEEE, 2018.
15. S. Ghosal, D. Blystone, A. K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, “An explainable deep machine vision framework for plant stress phenotyping,” Proceedings of the National Academy of Sciences, vol. 115, no. 18, pp. 4613-4618, 2018.
16. L. A. Wulandhari, A. A. S. Gunawan, A. Qurania, P. Harsani, T. Ferdy Tarawan, and R. F. Hermawan, “Plant nutrient deficiency detection using deep convolutional neural network,” ICIC Express Letters, vol. 13, no. 10, pp. 971-977, 2019.

17. J. D. Espineli and K. P. Lewis, “Internet of Things (IoT) based plant monitoring using machine learning,” in Advances in Information and Communication: Proc. Future of Information and Communication Conference (FICC), vol. 1, pp. 278–289, Springer, 2021.
18. Z. Xu, X. Guo, A. Zhu, X. He, X. Zhao, Y. Han, and R. Subedi, “Image-based diagnosis of nutrient deficiencies in rice using deep convolutional neural networks,” 2020.
19. B. S. Anami, N. N. Malvade, and S. Palaiah, “Deep learning approach for recognition and classification of yield affecting paddy crop stresses using field images,” Artificial Intelligence in Agriculture, vol. 4, pp. 12–20, 2020.
20. P. K. Sethy, N. K. Barpanda, A. K. Rath, and S. K. Behera, “Deep feature based rice leaf disease identification using support vector machine,” Computers and Electronics in Agriculture, vol. 175, p. 105527, 2020.
21. M. Sharma, K. Nath, R. K. Sharma, C. J. Kumar, and A. Chaudhary, “Ensemble averaging of transfer learning models for identification of nutritional deficiency in rice plant,” Electronics, vol. 11, no. 1, p. 148, 2022.
22. K. C. N. Jayasiri, S. Chandrasiri, S. Rupasinghe, K. O. R. Karunanayake, A. Uthayachandran, and M. I. F. Zihara, “Deep learning-based image analysis for detecting nutrient deficiencies of tomato plants,” in Proc. Int. Conf. Advancements in Computing (ICAC), pp. 209-214, IEEE, 2023.
23. A. Sathyan and P. Palanisamy, “CNN driven nutrient deficiency detection in plants using real-world leaf images,” Edelweiss Applied Science and Technology, vol. 8, no. 4, pp. 1483-1495, 2024.
24. J. Mkhatshwa, T. Kavu, and O. Daramola, “Analysing the performance and interpretability of CNN-based architectures for plant nutrient deficiency identification,” Computation, vol. 12, no. 6, p. 113, 2024.
25. S. Supreetha, R. Premalathamma, and M. S. H. Manjula, “Deep learning techniques to detect nutrient deficiency in rice plants,” in Proc. Int. Conf. Inventive Computation Technologies (ICICT), pp. 699-705, IEEE, 2024.
26. H. N. Ngugi, A. E. Ezugwu, A. A. Akinyelu, and L. Abualigah, “Revolutionizing crop disease detection with computational deep learning: A comprehensive review,” Environmental Monitoring and Assessment, vol. 196, no. 3, p. 302, 2024.
27. X. Jin, Y. Xie, X.-S. Wei, B.-R. Zhao, Z.-M. Chen, and X. Tan, “Delving deep into spatial pooling for squeeze-and-excitation networks,” Pattern Recognition, vol. 121, p. 108159, 2022.
28. S. M. Abbas and S. N. Singh, “Region-based object detection and classification using Faster R-CNN,” in Proc. 4th Int. Conf. Computational Intelligence & Communication Technology (CICT), pp. 1–6, IEEE, 2018.
29. S. R. Waheed, N. M. Suaib, M. S. M. Rahim, M. M. Adnan, and A. A. Salim, “Deep learning algorithms-based object detection and localization revisited,” Journal of Physics: Conference Series, vol. 1892, p. 012001, IOP Publishing, 2021.
30. P. Sunitha, “Images of nutrient deficient banana plant leaves,” Mendeley Data, vol. 1, 2022.