Journal of Theoretical and Applied Information Technology
30th September 2024. Vol.102. No. 18

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

COMPOSITE CROSS ATTENTION NETWORK FOR RELIABLE AND ROBUST PLANT NUTRITIONAL DEFICIENCY ANALYSIS: IMAGE ENHANCEMENT AND TRANSFER LEARNING

MRS. NIKITHA S[1], DR. PRABHANJAN S[2]

[1]Assistant Professor, Jyothy Institute of Technology, Department of Computer Science and Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, India
[2]Professor, Jyothy Institute of Technology, Department of Computer Science and Engineering, Affiliated to Visvesvaraya Technological University, Belagavi, India

E-mail: [1]nikitha.s@jyothyit.ac.in, [2]prabhanjan.s@jyothyit.ac.in

ID 55330 Submission	Editorial Screening	Conditional Acceptance	Final Revision Acceptance
14-08-24	16-08-2024	13-09-2024	25-09-2024

ABSTRACT

Nutritional deficiency analysis is crucial to enhance agricultural productivity and promote environmental sustainability. This study investigates domain-specific preprocessing techniques to improve the visual features of leaf images. Our proposed image enhancement pipeline uses edge enhancement filters, multiscale Contrast Limited Adaptive Histogram Equalization (CLAHE), and multiscale Retinex to improve the structural, textural, and colour details. "Composite Cross-Attention Network" is proposed to effectively integrate information from both primary and enhanced images. The image features were effectively extracted and integrated using transfer learning and a self-residual cross-attention block. The model achieved promising results with a 5-fold cross-validation strategy on the Mulberry and Rice datasets, achieving accuracies of 92.53% and 97.23%, respectively. It demonstrates superiority over models relying solely on primary images and shows robustness in real-time evaluations under varied environmental conditions. This study underscores the importance of advanced image processing and deep learning integration for optimizing nutritional deficiency analysis in precision agriculture, contributing to sustainable agricultural practices.
Keywords: *CLAHE, Retinex, DenseNet121, MobileNet, Cross-Attention*

1. INTRODUCTION

The precise management of nutritional deficiencies in agricultural practices is essential for sustaining agricultural productivity. The overapplication of fertilizers adversely affects water quality, soil health, and biodiversity [1]. Precise nutrient management practices promote sustainable agricultural productivity, minimize environmental impacts, and ensure long-term soil fertility [1].

Convolutional Neural Network (CNN) automates the process of learning features from data through hierarchical layers that progressively extract and abstract visual patterns [2]. In agricultural applications, CNN automates various tasks, such as plant disease detection, yield prediction, crop monitoring, and nutritional deficiency analysis, by efficiently processing and analyzing image datasets [3]. CNNs effectively analyze leaf images to identify signs of nutritional deficiencies such as discolouration, texture changes, and anomalies in leaf structure.

One of the primary challenges in analyzing nutritional deficiency in plants using CNN is the limited availability of labelled datasets, which hinders the effective training of deep neural networks. Training deep networks on small datasets often leads to overfitting and a lack of generalizability. Transfer learning mitigates these problems by leveraging pre-trained models trained on large-scale datasets, such as ImageNet [4]. Transfer learning enhances the generalization and performance of CNN in tasks, such as leaf image analysis, by transferring knowledge from these models.

Enhancements in image quality, such as improving contrast and colour information [5], reducing noise, and correcting illumination variations, are crucial for enhancing the interpretability and discriminative power of features extracted from leaf images in nutritional deficiency analysis. More precise and distinguishable visual details enable deep-learning models to identify subtle signs of plant nutritional deficiency stress, including discolouration, texture changes, and structural anomalies. Enhanced image quality ensures reliable and precise assessments of plant stress by mitigating the effects of poor lighting, shadows, and other distortions that obscure these critical features.

Based on domain-specific insights, this study proposes an image enhancement pipeline for leaf images to enhance their structural, textural, and colour properties. To effectively integrate information from both primary and enhanced images, we propose a Composite Cross-Attention Network (CCAN). Features are extracted from images using pre-trained models, such as DenseNet121 and MobileNet, and effectively integrated using a self-residual cross-attention block. The combined features are further processed through Spatial Pyramid Pooling (SPP), enabling detailed feature representation across multiple scales. The proposed approach significantly enhances the performance of nutritional deficiency analysis compared to models that rely only on primary images. In addition to performance improvement, the proposed approach is robust to environmental variations such as changes in lighting conditions.

Key Contributions:

- The proposed image enhancement pipeline leverages domain-specific insights to enhance the visual features of leaf images.
- Proposed Composite Cross-Attention Network (CCAN) for comprehensive plant nutritional deficiency analysis.

The remainder of this paper is organized as follows: Section 2 reviews related works. Section 3 describes the materials and methods. Section 4 presents the results and discussion. Finally, the paper concludes with a summary of key findings and further research directions.

2. RELATED WORKS

2.1 Transfer Learning Models for Nutritional Deficiency Analysis

Kolhar et al. [6] investigated deep neural networks like Xception, vision transformer, and MLP mixer models for classifying nutritional deficiencies in rice plants. The Xception model demonstrates superior accuracy and efficiency. Muthusamy et al. [7] accurately identified nutritional deficiencies in banana crops using ensembled InceptionV3 and DenseNet169 models. The ensemble model showed high accuracy, highlighting the effectiveness of transfer learning in adapting pre-trained models for agricultural applications. Venkatesh et al. [8] proposed a lightweight ensemble transfer learning model combining MobileNetV2 and shallow CNN architectures. The proposed technique attained higher accuracy in identifying nutritional deficiencies in rice and groundnuts. Shadrach et al. [9] proposed a hybrid approach integrating Gabor filters, Adam optimizer with SqueezeNet, and Ring Toss Game Optimization (RTGO) algorithm for detecting nutritional deficiencies in ridge gourds, showing high accuracy and specificity. Gul et al. [10] applied pre-trained CNN models, such as DenseNet and VGG16, to detect nutritional deficiencies in hydroponic basil, highlighting the importance of dataset diversity and model optimization for effective classification. Ramos-Ospina et al. [11] employed transfer learning with models such as VGG16, ResNet50, GoogLeNet, DenseNet201, and MobileNetV2 to detect phosphorus deficiency in maize, achieving high accuracy suitable for real-time crop monitoring. Nayak et al. [12] used smartphone images to classify 12 diseases and nutritional deficiencies in rice plants using DenseNet201, Xception, MobileNetV2, and ResNet50, emphasizing real-time disease detection capabilities, particularly with the MobileNetV2 model. Talukder et al. [13] developed a robust Deep Ensemble Convolutional Neural Network (DECNN) combining models such as InceptionV3, InceptionResNetV2, DenseNet121, DenseNet169, and DenseNet201 to diagnose rice nutritional deficiencies with high accuracy. Sharma et al. [14] introduced the DeepBatch framework and utilized models like Dilated U-EfficientNet and SegNet-EfficientNet for automated segmentation and classification of rice diseases and deficiencies, enhancing diagnostic precision. Sharma et al. [15] Leveraged cloud-based systems and six Transfer Learning (TL) models were used to enhance nutritional deficiency identification in rice,

emphasizing the role of ensemble approaches in optimizing crop yield through precise diagnosis.

Recent studies have utilized a transfer learning model to mitigate the challenge of limited labelled datasets for analyzing plant nutritional deficiencies. Models such as Xception, InceptionV3, DenseNet, and MobileNetV2 have been adapted and fine-tuned using pre-trained weights. Transfer learning enables these models to extract high-level features relevant to plant characteristics. By initializing with pre-trained weights, the models start with parameters that have already learned the hierarchical representations of the features. This initialization often accelerates the training process, allowing the model to converge faster and achieve higher accuracy with fewer labelled examples. Thus, leveraging transfer learning optimizes limited labelled datasets, enhances feature extraction capabilities, and accelerates training. Careful fine-tuning of these models is essential for mitigating overfitting and ensuring robust performance across diverse agricultural conditions.

2.2 Image Enhancement Techniques for Improved Accuracy in Deep Learning

Rifai et al. [16] applied white balance correction and CLAHE to significantly improve the performance of the MobileNetV2 model for analyzing chest X-ray images, emphasizing enhanced image clarity and diagnostic accuracy. Hayati et al. [17] applied CLAHE to improve contrast and image quality in diabetic retinopathy classification with EfficientNetB4, showcasing its effectiveness in enhancing model accuracy. Hassan et al. [18] utilized Histogram Equalization (HE) and CLAHE for preprocessing in gender classification, achieving notable accuracy improvements on datasets LFW (Labeled Faces in the Wild). Saifullah et al. [19] integrated CLAHE-HE preprocessing with a CNN-U-Net model to enhance brain tumour segmentation accuracy by improving image contrast and quality, which is crucial for precise image analysis tasks. Rasheed et al. [20] explored Gaussian blur and CLAHE techniques to enhance MRI images for brain tumour classification, demonstrating significant accuracy gains through improved feature extraction capabilities of CNNs. Macsik et al. [21] employed RGB-CLAHE and Lab-CLAHE to enhance fundus images for diabetic retinopathy classification, focusing on contrast enhancement and noise reduction to improve deep-learning model performance. Sundar et al. [22] applied CLAHE to enhance image contrast and quality in histopathology images, achieving exceptional accuracy in tissue classification using ResNet50V2 and ensemble learning. Mahmood et al. [23] investigated Multi-Scale Retinex Guided Filter (MSR-GF) for enhancing image quality in surveillance applications, addressing challenges like non-uniform illumination to improve image clarity and detail. Agrawal et al. [24] highlighted CLAHE's role in enhancing mammogram image quality for classification tasks, ensuring better visibility of structural information essential for accurate analysis using CNNs and DCNNs. Prince et al. [25] applied CLAHE-YCrCb and LBP techniques to enhance COVID-19 detection from chest X-ray images, leveraging chroma components and local contrast improvements for effective disease identification. Patel et al. [26] applied CLAHE for contrast enhancement to improve the visibility of subtle pathological features in Ulcerative Colitis images. This technique enhances the overall quality and clarity of the images, facilitating better feature extraction. Using CLAHE combined with data augmentation significantly improved the classification accuracy of MobileNet-V2.

Advanced image enhancement techniques, such as CLAHE and Retinex, have significantly improved the performance of deep learning models in medical and surveillance applications. CLAHE, as demonstrated in medical imaging research [16-22], [24-26], enhances the local contrast and image quality to highlight subtle details. Similarly, Retinex techniques, such as MSR-GF, applied in surveillance contexts [23], focus on improving image clarity and detail, facilitating precise feature extraction for classification tasks.

In the context of plant nutritional deficiency analysis, studies have highlighted the challenges of limited labelled datasets and environmental variations. To address these issues, we propose an image enhancement pipeline that utilizes edge enhancement filters, CLAHE, and Retinex techniques to improve leaf image's structural, textural, and colour information. This preprocessing enhances image details, facilitating more accurate analyses. Additionally, we incorporate transfer learning using pre-trained models like DenseNet121 and MobileNet to extract high-level features. This approach accelerates the training process and improves model performance, maximize the utility of limited labelled datasets while ensuring consistent results across diverse agricultural conditions.

3. MATERIALS AND METHODS

3.1 Data Acquisition and Preprocessing

Experiments were conducted in a controlled environment to develop visual symptoms of nutritional deficiencies in mulberry plants. Healthy samples were grown using adequate fertilization. In addition to the experimental samples, mulberry leaves exhibiting nutritional deficiencies were collected from mulberry fields in Kanakapura taluk and Ramanagara districts in Karnataka, India. The dataset included both experimental and field images [27]. Healthy Mulberry leaves are characterized by their thickness, broadness, and rich green pigmentation. In contrast, leaves affected by nutritional deficiencies display symptoms such as uniform chlorosis, unusual dark green colour, brown lesions and interveinal chlorosis [28].

Leaf image data were categorized into six classes based on symptoms of nutritional deficiency. The distribution of images across each category is as follows: Healthy -227, Nitrogen (N)- 131, Phosphorus (P) -100, Potassium (K) -180, Sulphur (S) -133 and Iron (Fe) -180. Representative samples from each category were subjected to leaf tissue analysis at the Indian Institute of Horticultural Research (ICAR) in Hesaraghatta, Bengaluru, to confirm the presence of nutritional deficiencies in mulberry leaves.

Images were captured using a 12 MP smartphone camera under controlled conditions with a white background, partial shading, and ambient light. A series of preprocessing steps - K-means clustering, morphological operations, and contour detection, was applied to isolate the leaf region by removing the background. The extracted leaf images were resized for uniformity, facilitating detailed analysis [27].

The proposed approach was also applied to the Rice [29] dataset from the Kaggle Public Repository. The distribution of images across each category is as follows: Nitrogen (N) -440, Phosphorus (P) -333 and Potassium (K) -383.

3.2 Leaf Image Enhancement Pipeline

Nutritional deficiencies in plants often manifest as changes in vein patterns, colouration, and textural properties. To enhance these critical indicators, a leaf image enhancement pipeline is proposed. This pipeline utilizes a Sobel filter [30] for edge enhancement, multiscale CLAHE [31] for texture enhancement and multiscale Retinex [32] for colour enhancement.

Proposed Leaf Image Enhancement Pipeline

Input: Preprocessed leaf image
Output: Enhanced leaf image

Step 1: Edge enhancement using the Sobel filter.

1. Colour Space Conversion: Convert the input image to LAB colour space.
2. Apply Sobel operator to Luminance (L) channel:
   a. Compute gradient

$$G_x = \frac{\partial L}{\partial x} \quad (1)$$

Equation (1) computes the gradient in the x direction.

$$G_y = \frac{\partial L}{\partial y} \quad (2)$$

Equation (2) computes the gradient in the y direction.

   b. Compute the gradient magnitude

$$G = \sqrt{G_x^2 + G_y^2} \quad (3)$$

Equation (3) computes the magnitude of the gradient.

3. Normalize the gradient magnitude to the range [0,255] and merge it with the A and B Channels.
4. Colour Space Conversion: Convert the LAB image to BGR colour space.
5. Blend the input with the Sobel image using a weighted approach to retain structural and colour information.

$$I_{Sobel\_blend} = 0.8 \times I_{input} + 0.2 \times I_{Sobel} \quad (4)$$

In equation (4), I_{Sobel_blend} represents the blended image, I_{input} represents the input image and I_{Sobel} represents the edge-enhanced image.

Step 2: Texture details enhancement using multiscale Contrast Limited Adaptive Histogram Equalization (CLAHE)

1. Colour Space Conversion: Convert the edge-enhanced image to LAB colour space.
2. Define grid sizes: Set the tile grid sizes $[(n_1, n_1), (n_2, n_2), \ldots, (n_k, n_k)]$. These grid sizes determine the size of local regions for contrast enhancement.
3. CLAHE for each grid size:
   a. Initialize L_{CLAHE}=0
   b. For each grid size (n_i, n_i):
      - Compute CLAHE on the L channel.
        - Calculate the histogram H_T of the L channel within the grid:

$$H'_T(i) = \min(H_T(i), M) + \frac{\sum_{j=0}^{255} \max(H_T(j) - M, 0)}{256} \quad (5)$$

In equation (5), $H'_T(i)$ represent the adjusted histogram value at intensity level i, $H_T(i)$ represents the original histogram value at intensity level i, M denotes the clip limit, and j is the index variable iterating over intensity levels.

- Compute the cumulative distribution function (CDF)

$$CDF_T(i) = \sum_{j=0}^{i} H'_T(j) \quad (6)$$

In equation (6), $CDF_T(i)$ represent CDF of the transformed histogram $H'_T(j)$ up to intensity level i and $H'_T(j)$ denotes the transformed histogram value at intensity level j obtained after applying CLAHE to the local regions of the image.

- Transform the L channel $L'_T(x, y)$

$$L'_T(x, y) = \frac{CDF_T(L_T(x, y)) - CDF_T(0)}{n_i^2 - CDF_T(0)} X\ 255 \quad (7)$$

In equation (7), $L'_T(x, y)$ represents the transformed luminance value at pixel (x, y) after CLAHE processing, $CDF_T(L_T(x, y))$ denotes the CDF of the transformed histogram $H'_T(j)$ which is evaluated at the luminance intensity $L_T(x, y)$, $CDF_T(0)$ signifies the CDF value at intensity level 0 of the transformed histogram $H'_T(j)$, providing a baseline for contrast enhancement by CLAHE and n_i indicating the size of the grid used for CLAHE processing, specifying the number of pixels in each dimension of the grid.

- Aggregate $L'_T(x, y)$ to L_{CLAHE}

$$L_{CLAHE} += L'_T(x, y) \quad (8)$$

4. Average the aggregated L_{CLAHE} results using the number of grid sizes.
5. Clip L_{CLAHE} to range [0,255] and merge L_{CLAHE} with A and B channels
6. Colour Space Conversion: Convert the LAB image to BGR colour space.

Step 3: Enhance colour details using multi-scale Retinex

1. Convert edge and texture-enhanced image to float 32 and normalize its intensity values to the range [0,1]
2. Apply multi-scale Retinex
   - Initialize the list of sigma values:

$$\sigma_{lis} = [small_{value}, medium_{value}, large \quad (9)$$

- Initialize an empty array for the Retinex-enhanced image

$$I_{Retinex} = log_{10}(I_{CLAHE} + \in) \quad (10)$$

In equation (10), $I_{Retinex}$ represents the Retinex-enhanced image, $log_{10}(I_{CLAHE} + \in)$ denotes the logarithm base 10 of the $I_{CLAHE} + \in$, ϵ is a small constant to avoid taking the logarithm of zero.

- for each σ in σ_{list}

$$I_{Retinex} += log_{10}(I_{CLAHE} + \in) - mean(log_{10}(GaussianBlur(I_{CLAHE} + \in, \sigma))) \quad (11)$$

In equation (11), $mean(log_{10}(GaussianBlur(I_{CLAHE} + \in, \sigma)))$ represents the mean value of the logarithm base 10 of the image $GaussianBlur(I_{CLAHE} + \in, \sigma)$, which is Gaussian-blurred using a kernel with a standard deviation σ and σ Standard deviation of the Gaussian kernel used for the blur operation. It controls the extent of smoothing applied to the image.

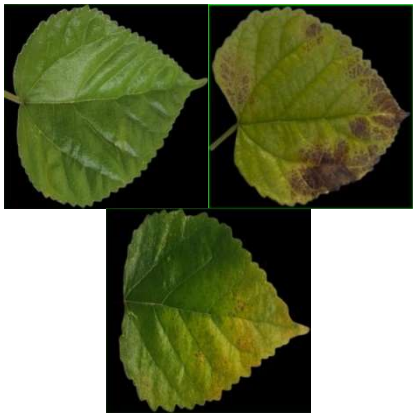
- Average the Retinex results and clip the values to ensure they are within the range [0,1]

3. Convert the image back to UInt8

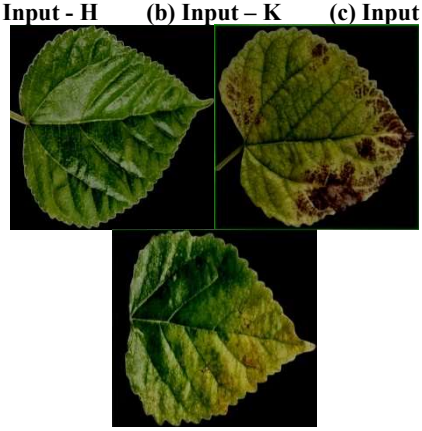
The leaf image enhancement pipeline uses an edge enhancement filter to enhance structural details. Initially, the image is converted to the LAB colour space to separate luminance (L) and colour information. The Sobel operator with a filter size of 3x3 is applied to the L channel to compute gradients in horizontal (Gx) and vertical (Gy) directions. These gradients are combined to calculate the gradient magnitude (G) normalized to a range of 0 to 255 to enhance edge clarity. This enhanced gradient, termed $Gnorm$, is integrated with the A and B colour channels within the LAB colour space. The resulting image is converted to the standard colour space (BGR) and blended with the original input using a weighted average to effectively preserve the structure and colour information.

Following edge enhancement, texture details are enhanced using multiscale CLAHE. Different grid sizes (4x4, 8x8, 16x16) are defined to enhance contrast locally within the L channel. CLAHE adjusts histograms to ensure optimal contrast without over-enhancement by applying a clip limit (3.0), which controls the maximum intensity adjustment and helps prevent excessive contrast and noise amplification. Normalized histograms are used to compute Cumulative Distribution Functions (CDFs), guiding the transformation of the L channel. These transformed values are aggregated and averaged to produce L_{CLAHE}, clipped to 0 to 255 to maintain image integrity. L_{CLAHE} is then merged with the A and B channels within the LAB colour space before converting back to the BGR for further processing.

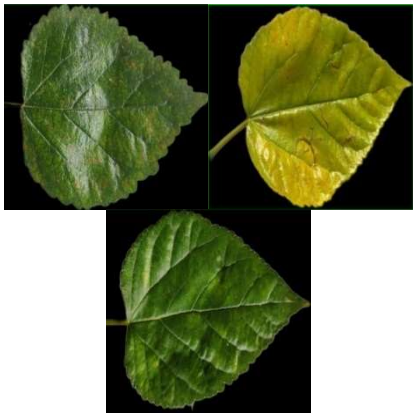
The final step focuses on enhancing colour details through multiscale Retinex processing. The CLAHE-enhanced image is converted to a float32 format, and its intensity values are normalized to a range of 0 to 1. The Retinex algorithm is applied using a series of sigma values (10,30,50), which control the scale of Gaussian blurring applied to the image. Logarithmic transformations are iteratively applied, and results are averaged to ensure consistent enhancement across different scales. After clipping intensity values to 0 to 1, the image is converted to UInt8 format. Enhanced leaf images preserve refined colour details and overall image quality. Figure 1 shows the leaf image enhancement pipeline's sample input and output images.



(a) Input - H (b) Input – K (c) Input – N



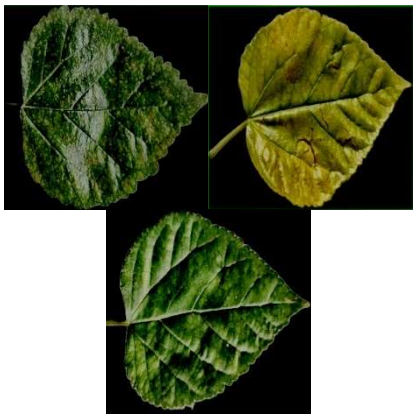
(a) Output - H (b) Output – K (c) Output – N



(d) Input – P (e) Input – S (f) Input - Fe

Journal of Theoretical and Applied Information Technology
30th September 2024. Vol.102. No. 18

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



(d) Output – P (e) Output – S (f) Output – Fe

Figure 1: Sample Input and Output Images

3.3 Composite Cross-Attention Network

The proposed model utilizes both primary and enhanced images to effectively address the artefacts introduced in the enhanced image while preserving the authenticity of the primary image. Features from this composite image are extracted using a pre-trained model known for its ability to capture hierarchical features. To integrate these features effectively, a self-residual cross-attention mechanism is employed. Additionally, Spatial Pyramid Pooling (SPP) extracts features from this concatenated image at multiple scales. Subsequently, the dense layers fine-tune these features, leading to the final output layer. Figure 2 illustrates the architecture of the proposed Composite Cross-Attention Network.

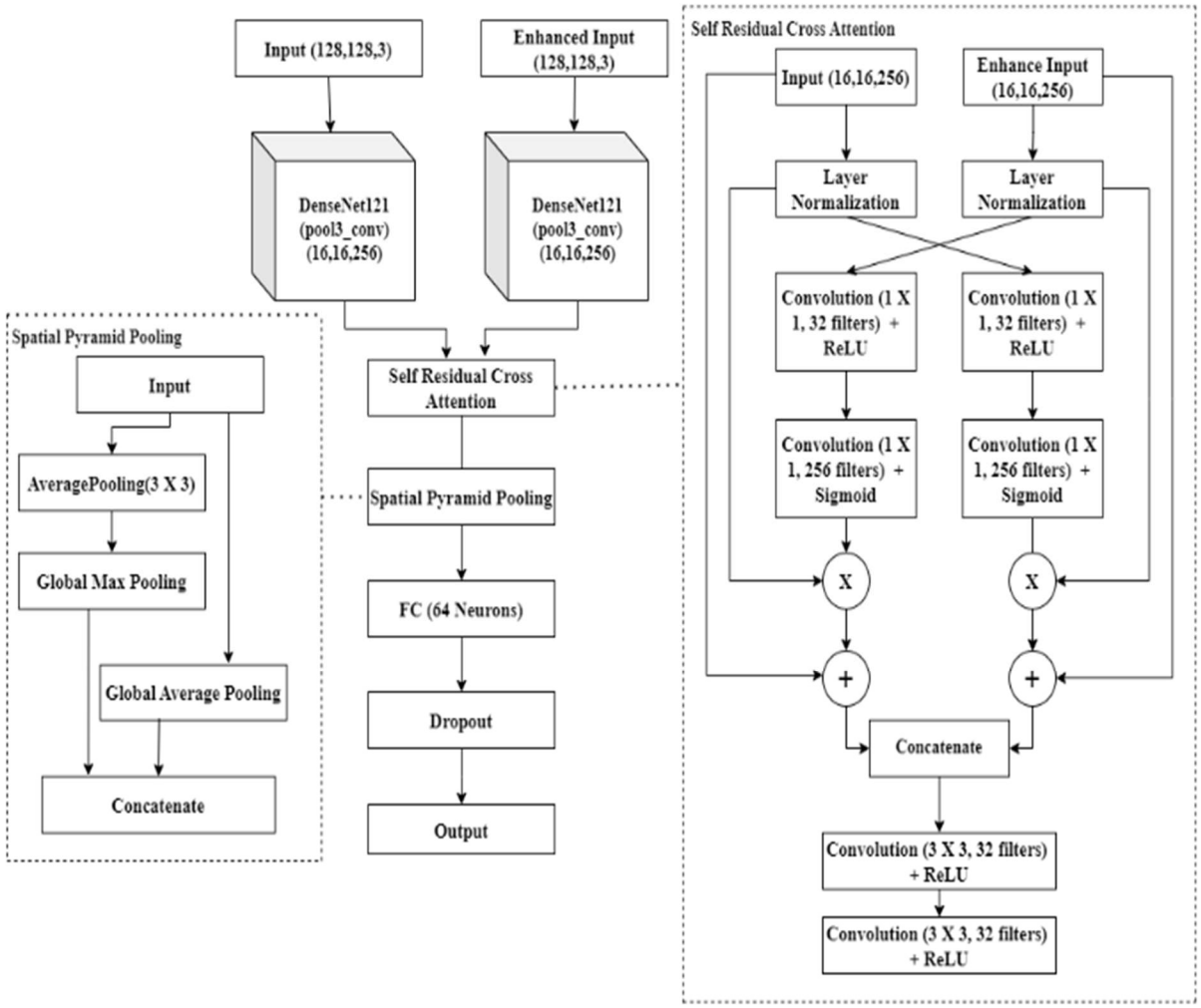


Figure 2: Composite Cross-Attention Network

3.3.1 Feature Extraction from Pre-trained Model

The proposed model utilizes the features extracted from a pre-trained model. The DenseNet121 [33] architecture was selected for feature extraction based on its performance on the dataset. Features were extracted from the intermediate layers of the pre-trained model to balance the rich hierarchical details and adequate spatial dimensions for further analysis. These intermediate representations capture critical hierarchical spatial features for a robust image analysis. DenseNet121 features from the 'pool3_conv' layer excel at capturing complex spatial patterns through hierarchical abstraction. By leveraging the outputs from this intermediate layer of the pre-trained DenseNet121 model, our method ensures a balanced capture of the detailed local features and broader global contexts required for accurate image analysis.

To demonstrate the model's adaptability to different pretrained models, it was tested with MobileNet [34] for feature extraction. The output from MobileNet's 'conv_pw_5_relu' intermediate layer was used for subsequent analysis. Experimenting with two distinct architectures for feature extraction showed the model's ability to adapt to various architectural frameworks.

3.3.2 Self-Residual Cross Attention Block

A self-residual cross-attention [35] block blends information from the primary and enhanced leaf images. This mechanism enhances informative regions by computing attention weights highlighting relevant spatial contexts while suppressing noise. Applied layer normalization to both the primary and enhanced features. Subsequently, attention weights are calculated for the primary features based on the enhanced features using convolutional layers with ReLU activation. Similarly, the attention weights for the enhanced features are computed based on the primary features. These computed attention weights adaptively adjust the features and merge them with the inputs using self-residual connections to preserve the essential spatial information. Finally, the combined features undergo further refinement through convolutional layers to effectively emphasize the critical image characteristics.

3.3.3 Pooling Operation and Dense Layer

In addition to feature refinement, our architecture incorporates the pooling operation to aggregate spatial information and a dense layer to enhance feature representation, followed by an output layer. Based on the architecture used for feature extraction, the specific pooling operation is selected to optimize performance. For DenseNet, Spatial Pyramid Pooling (SPP) [27],[36] effectively captures and integrates spatial information at various scales, enhancing feature representation. In contrast, for MobileNet, Global Average Pooling (GAP) is employed instead of SPP to align with its lightweight design and computational efficiency. Following the pooling operation, a fully connected layer with 64 neurons is used to refine the features before passing them to the final output layer.

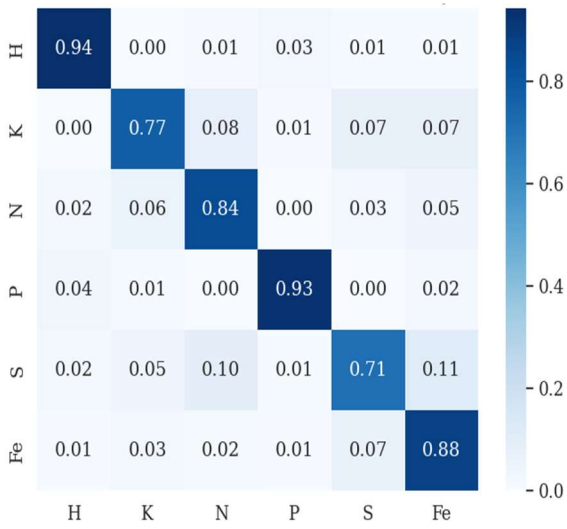
Early stopping and adaptive learning rate strategies mitigate overfitting during training. Early stopping is configured using a 12-epoch patience threshold. The ReduceLROnPlateau callback reduces the learning rate by 0.1 after five epochs without improvement, with a minimum rate of 1e-5. The Nadam optimizer is used with a base learning rate of 0.0001. Together, these hyperparameters effectively optimize the training process.

4. RESULTS AND DISCUSSION

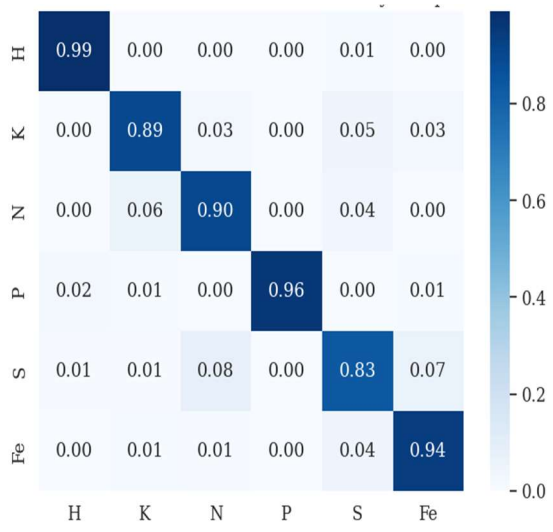
The models are implemented using the TensorFlow framework on the Google Colab Pro+ platform. The proposed CCAN is tested on the Mulberry and Rice datasets using a 5-fold cross-validation strategy. This approach involved training four folds and validating on the fifth fold, aggregating the results to ensure robustness. To address the class imbalance, the synthetic minority oversampling technique (SMOTE) is employed. Two diverse pretrained models, DenseNet121 and MobileNet, were tested for feature extraction from intermediate layers ('pool3_conv' for DenseNet121 and 'conv_pw_5_relu' for MobileNet) to demonstrate the adaptability of the proposed approach on different architectures.

The confusion matrix summarizes the performance of the classification model. It shows the number of correct and incorrect predictions made by the model across all the classes. This matrix effectively evaluates the model's accuracy, precision, recall, and F1 score for each class, providing insights into its overall performance and

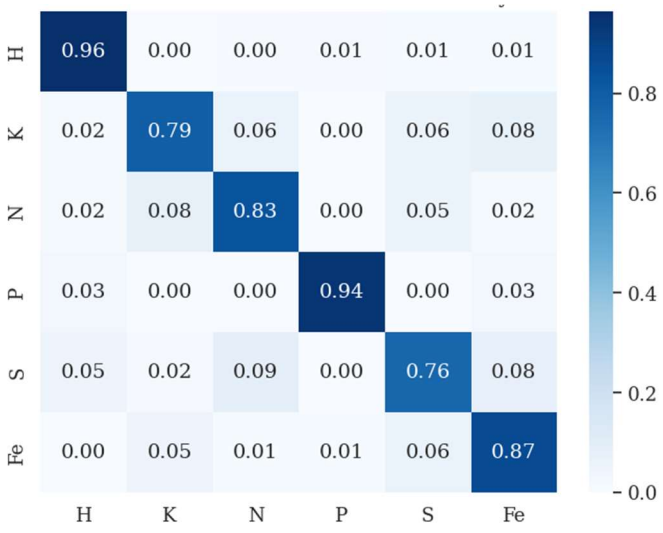
areas for improvement in classification tasks. Figure 3 compares the normalized confusion matrices for the mulberry dataset using our proposed CCAN versus models relying only on primary images for feature extraction. For the model relying only on primary images, features were extracted from the base model, followed by global average pooling and a fully connected layer of 64 neurons. Table 1 presents the class-wise precision, recall, and F1 score.



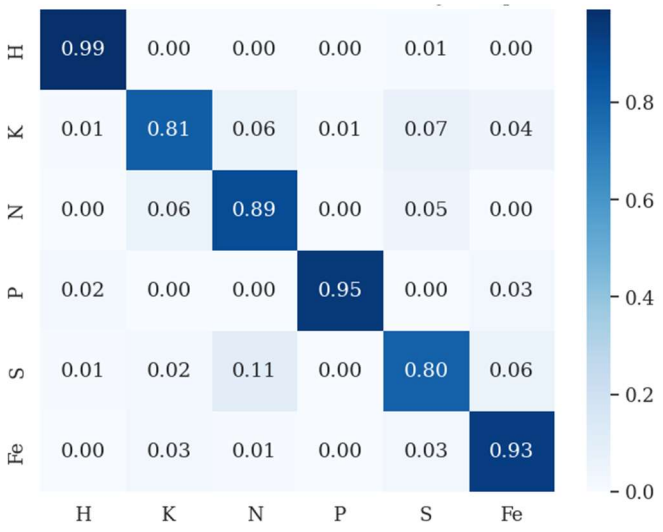
(a) Base Model (DenseNet121) – Primary Image



(b) CCAN (DenseNet121) - Primary and Enhanced Image



(c) Base Model (MobileNet) – Primary Image



(d) CCAN (MobileNet) – Primary and Enhanced Image

Figure 3: Normalized confusion matrix comparison - evaluated on the Mulberry Dataset.

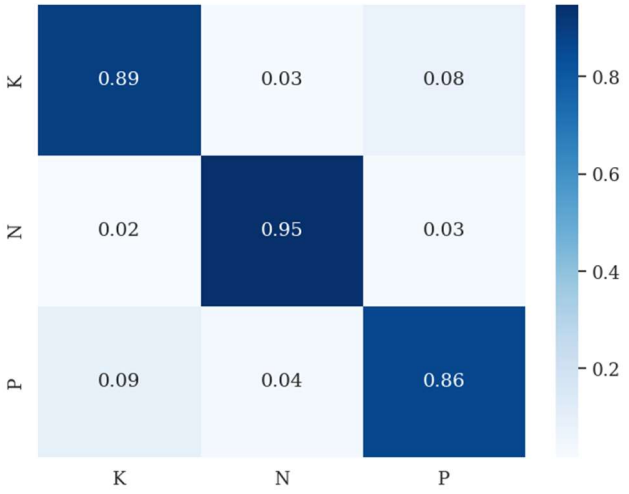
Table 1: Class-wise performance - evaluated on the Mulberry dataset.

Model	Class	Precision	Recall	F1 Score	Accuracy
Base Model (DenseNet121) -	H	0.9511	0.9427	0.9469	0.8505
	K	0.8742	0.7722	0.8201	

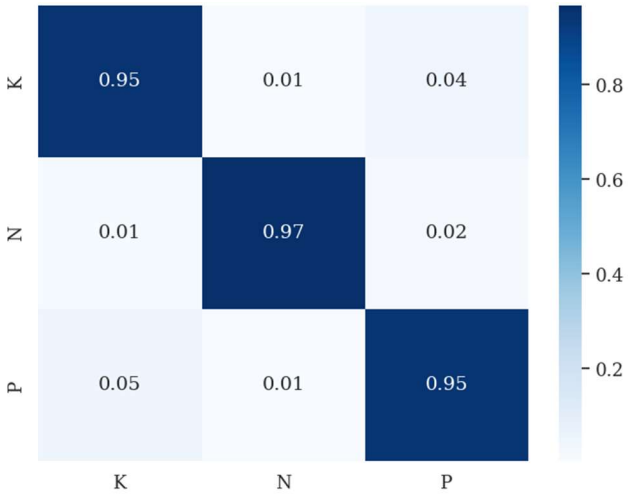
Primary Image	N	0.7692	0.8397	0.8029	
	P	0.9109	0.9293	0.9200	
	S	0.7540	0.7143	0.7336	
	Fe	0.8061	0.8778	0.8404	
CCAN (DenseNet121) – Primary and Enhanced images	H	0.9868	0.9912	0.9890	0.9253
	K	0.9306	0.8944	0.9122	
	N	0.8741	0.9008	0.8872	
	P	1.0000	0.9596	0.9794	
	S	0.8222	0.8346	0.8284	
	Fe	0.9185	0.9389	0.9286	
Base Model (MobileNet) -Primary Image	H	0.9359	0.9648	0.9501	0.8632
	K	0.8659	0.7889	0.8256	
	N	0.8074	0.8321	0.8195	
	P	0.9490	0.9394	0.9442	
	S	0.7710	0.7594	0.7652	
	Fe	0.8298	0.8667	0.8478	
CCAN (MobileNet) – Primary and Enhanced images	H	0.9825	0.9912	0.9868	90.09
	K	0.9006	0.8101	0.8529	
	N	0.8227	0.8855	0.8529	
	P	0.9792	0.9495	0.9641	
	S	0.7926	0.8045	0.7985	
	Fe	0.8984	0.9333	0.9155	

The normalized confusion matrix and Table 1 class-wise results show that the proposed CCAN using composite images for feature extraction significantly improved performance compared to the model relying solely on the primary image. CCAN enhanced performance across all classes; however, the models misclassified Potassium, Nitrogen, Sulphur, and iron samples. This can be attributed to overlapping visual symptom similarities between these classes. Further studies could explore effective preprocessing techniques to address these challenges.

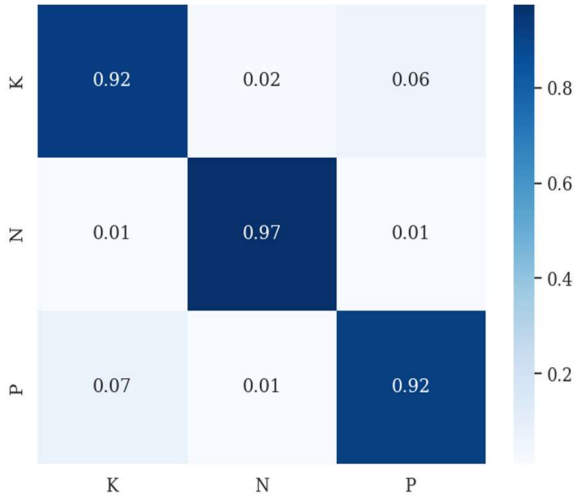
Figure 4 and Table 2 show the proposed CCAN's performance on the Rice dataset. The normalized confusion matrix and class-wise results showed that CCAN using composite images performed better on all classes than the model relying only on primary images for feature extraction.



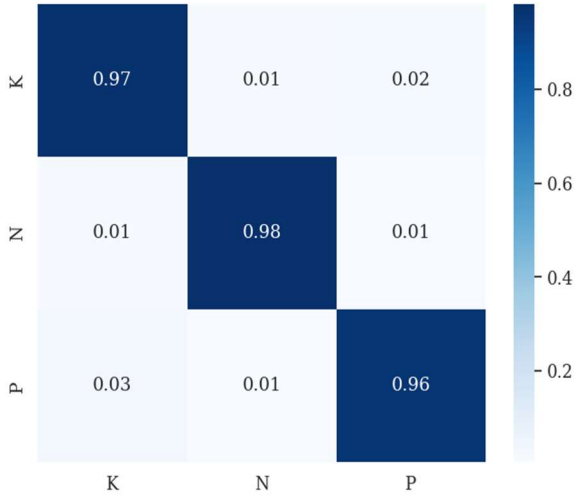
(a) Base Model (DenseNet121) – Primary image



(b) CCAN (DenseNet121) - Primary and Enhanced Image



(C) Base Model (Mobilenet) – Primary Image



(D) CCAN (Mobilenet) – Primary And Enhanced Image

Figure 4: Normalized Confusion Matrix Comparison - Evaluated On The Rice Dataset.

Table 2: Class-Wise Performance - Evaluated On The Rice Dataset.

Model	Class	Precision	Recall	F1 Score	Accuracy
Base Model (DenseNet121) - Primary Image	K	0.8974	0.8903	0.8938	0.9048
	N	0.9434	0.9477	0.9456	
	P	0.8623	0.8649	0.8636	
CCAN (DenseNet121) – Primary and Enhanced images	K	0.9907	0.9682	0.9793	0.9394
	N	0.9292	0.9459	0.9375	
Base Model (MobileNet) -Primary Image	K	0.9239	0.9191	0.9215	
	N	0.9750	0.9750	0.9750	
	P	0.9104	0.9159	0.9132	
CCAN (MobileNet) – Primary and Enhanced images	K	0.9588	0.9713	0.9650	0.9723
	N	0.9841	0.9818	0.9829	
	P	0.9726	0.9610	0.9668	

Table 3 presents the overall performance of both approaches on the mulberry dataset, and Figure 5 presents a comparison graph. The CCAN, using composite images and leveraging the DenseNet121 model for feature extraction, achieved superior performance on the Mulberry dataset.

Table 3: Overall Performance - Evaluated On The Mulberry Dataset.

Model	Precision	Recall	F1 Score	Accuracy
Base Model (DenseNet121) - Primary Image	0.8443	0.8460	0.8440	0.8505
CCAN (DenseNet121) – Primary and Enhanced images	0.9220	0.9199	0.9208	0.9253
Base Model (MobileNet) - Primary Image	0.8598	0.8585	0.8587	0.8632
CCAN (MobileNet) – Primary and Enhanced images	0.8960	0.8957	0.8951	0.9009

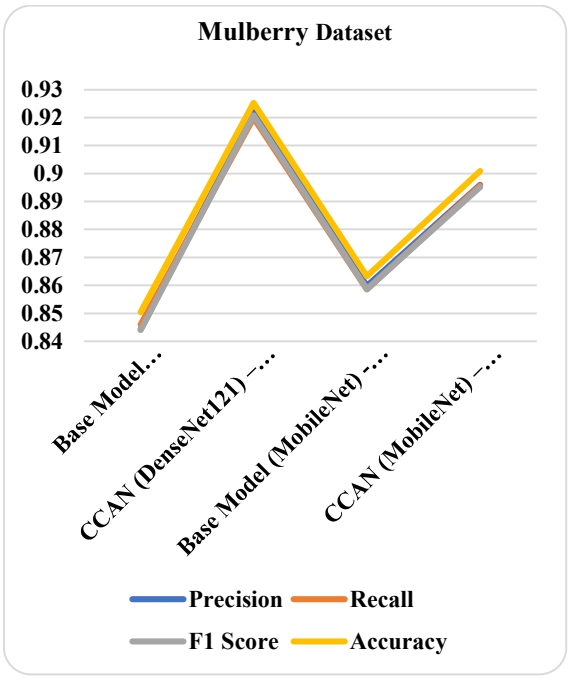


Figure 5: Graph Showing Performance Comparison – Evaluated On The Mulberry Dataset.

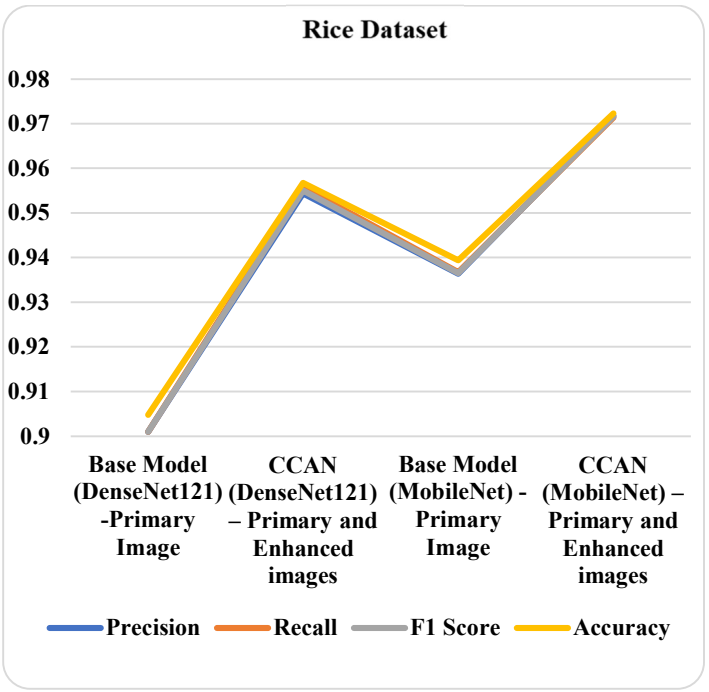


Figure 6: Graph Showing Performance Comparison - Evaluated On The Rice Dataset.

Table 4 presents the overall performance, and Figure 6 shows a comparison graph for the Rice dataset. The CCAN achieved superior performance in the Rice dataset by using composite images and leveraging the MobileNet model for feature extraction.

Table 4: Overall Performance – Evaluated On The Rice Dataset

Model	Precision	Recall	F1 Score	Accuracy
Base Model (DenseNet121) -Primary Image	0.9010	0.9010	0.9010	0.9048
CCAN (DenseNet121) – Primary and Enhanced images	0.9544	0.9557	0.9550	0.9567
Base Model (MobileNet) - Primary Image	0.9364	0.9367	0.9365	0.9394
CCAN (MobileNet) – Primary and Enhanced images	0.9718	0.9714	0.9716	0.9723

Experiments on both datasets show that CCAN using the primary and enhanced images rich in structural, texture, and colour details significantly improves performance. However, there were misclassifications between Nitrogen, Potassium, Sulphur, and iron samples in the Mulberry dataset. Severe Nitrogen-deficient samples were misclassified as Potassium-deficient, while early symptoms of Potassium were misclassified as Nitrogen-deficient. This issue can be addressed by collecting more samples that exhibit subtle differences and segregating the images based on severity levels.

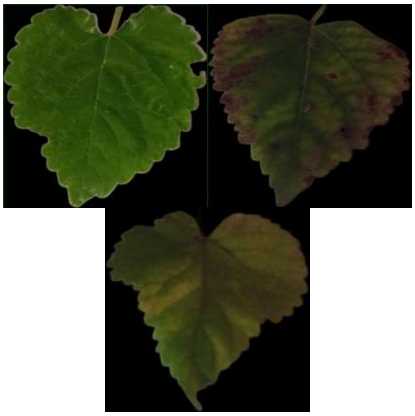
In advanced stages, Sulphur deficiency can lead to intervein chlorosis, similar to Iron deficiency, but veins are lighter in colour and thinner. Vein structure is crucial for distinguishing them. A better approach, like vein annotation, can be explored further to emphasize vein thickness and colour rather than relying on classical filters for edge enhancement in the image enhancement pipeline.

Furthermore, Nitrogen and sulfur deficiencies exhibit similar visual symptoms of chlorosis, but sulfur deficiency typically affects young leaves first, while nitrogen deficiency affects older leaves. A combined approach, considering

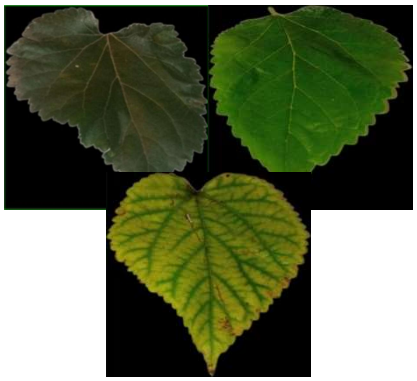
both old and young leaves in decision-making, may be effective.

We also tested using only enhanced images to extract features, but the improvement was marginal because of the artefacts. Combining the enhanced image with the primary image significantly improves performance. This approach capitalizes on the enhanced image's enriched visual characteristics and the original image's stability and authenticity, ensuring a more robust and accurate feature extraction process. Thus, integrating primary and enhanced images boosts performance and preserves critical image attributes essential for reliable feature extraction and analysis.

Furthermore, we tested the CCAN on real-time data. The CCAN model generalized well on real-time data and was robust to environmental changes, such as shadows and lighting effects. We collected real-time leaf samples under varying conditions, and our proposed approach demonstrated robustness to these changes. In contrast, the base model trained only on primary images experienced performance degradation. Figure 7 shows sample real-time test images, and Table 5 summarizes the overall precision, recall, F1 score, and accuracy.



(a) H N **(b) K** **(c)**



(D) P *(E) S* *(F) Fe*

Figure 7: Real-Time Test Samples

Table 5: Performance Comparison On Real-Time Data – Evaluated On The Mulberry

Model	Precision	Recall	F1 Score	Accuracy
Base Model (DenseNet121) -Primary Image	0.8223	0.8645	0.8154	0.8025
CCAN (DenseNet121) – Primary and Enhanced Images	0.8958	0.9022	0.8922	0.9136
Base Model (MobileNet) -Primary Image	0.8056	0.8177	0.7838	0.8025
CCAN (MobileNet) – Primary and Enhanced Images	0.8608	0.8836	0.8680	0.8889

The proposed CCAN is lightweight in terms of the number of trainable parameters and the training time. The total number of trainable parameters is 195654. The model converged within 50 epochs for both datasets. The training used 5.4 GB of RAM and a GPU with 1.2 GB of memory.

Our study demonstrated the impact of effectively integrating both primary and enhanced images using a cross-attention mechanism. The combined approach preserves fine-grained details and improves visual clarity. Our approach significantly enhances the overall performance of the nutritional deficiency analysis compared to those

relying solely on primary images. This dual-image strategy effectively mitigates the challenges posed by environmental variations, such as lighting variations, where models trained only on primary images exhibit degraded performance due to reduced visual distinctiveness.

Further studies could explore extending the dataset to include minute details through controlled experiments or using generative adversarial networks to generate synthetic images to mitigate misclassification between Potassium and nitrogen. Advanced leaf vein annotation techniques could be investigated to address misclassification between Sulphur and iron. Old and young leaves can be considered for decision-making to address misclassification between Nitrogen and Sulfur.

5. CONCLUSION

Our proposed CCAN, which combines primary with enhanced images enriched with structural, textural, and colour details, outperformed models relying only on primary images. This strategy enhanced the precision, recall, F1 score, and overall accuracy of the Mulberry and Rice datasets and ensured robustness against environmental variations. Our experiments underscore the importance of incorporating advanced image processing with deep learning techniques to improve the robustness and reliability of nutrition deficiency analysis.

Further research could enhance the datasets by incorporating minute details to address misclassifications caused by visual similarities. Additionally, it could explore advanced annotation techniques to improve vein feature representation and integrate both old and young leaves into the decision-making process. Integrating primary and enhanced images presents a promising avenue for advancing agricultural diagnostics, enabling reliable plant nutritional deficiency analysis across various environmental conditions.

6. ACKNOWLEDGEMENT

This work is supported by Visvesvaraya Technological University under VTU research grant scheme 2021.

REFERENCES

[1] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, “Global food demand and the sustainable intensification of agriculture,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 50, pp. 20260–20264, Dec. 2011, doi: 10.1073/pnas.1116437108.

[2] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[3] A. Kamilaris and F. X. Prenafeta-Boldú, “A review of the use of convolutional neural networks in agriculture,” *J. Agric. Sci.*, vol. 156, no. 3, pp. 312–322, Apr. 2018, doi: 10.1017/S0021859618000436.

[4] S. J. Pan and Q. Yang, “A Survey on Transfer Learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[5] Y. Qi *et al.*, “A Comprehensive Overview of Image Enhancement Techniques,” *Arch. Comput. Methods Eng.*, vol. 29, no. 1, pp. 583–607, Jan. 2022, doi: 10.1007/s11831-021-09587-6.

[6] S. Kolhar, J. Jagtap, and R. Shastri, “Deep Neural Networks for Classifying Nutrient Deficiencies in Rice Plants Using Leaf Images”.

[7] 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, doi: 10.1109/ACCESS.2024.3405027.

[8] K. Venkatesh and K. J. Naik, “An ensemble transfer learning for nutrient deficiency identification and yield-loss prediction in crop,” *Multimed. Tools Appl.*, Feb. 2024, doi: 10.1007/s11042-024-18592-3.

[9] “Optimal transfer learning based nutrient deficiency classification model in ridge gourd (Luffa acutangula) | Scientific Reports.” Accessed: Jul. 25, 2024. [Online]. Available: https://www.nature.com/articles/s41598-023-41120-6

[10] “Sensors | Free Full-Text | Exploiting Pre-Trained Convolutional Neural Networks for the Detection of Nutrient Deficiencies in Hydroponic Basil.” Accessed: Jul. 25, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/23/12/5407

[11] “Electronics | Free Full-Text | Deep Transfer Learning for Image Classification of Phosphorus Nutrition States in Individual Maize Leaves.” Accessed: Jul. 25, 2024. [Online].

Available: https://www.mdpi.com/2079-9292/13/1/16
[12] "Application of smartphone-image processing and transfer learning for rice disease and nutrient deficiency detection - ScienceDirect." Accessed: Jul. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375523000254
[13] "Nutrients deficiency diagnosis of rice crop by weighted average ensemble learning - ScienceDirect." Accessed: Jul. 25, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772375522001198
[14] M. Sharma, C. J. Kumar, J. Talukdar, T. P. Singh, G. Dhiman, and A. Sharma, "Identification of rice leaf diseases and deficiency disorders using a novel DeepBatch technique," *Open Life Sci.*, vol. 18, no. 1, Jan. 2023, doi: 10.1515/biol-2022-0689.
[15] "Electronics | Free Full-Text | Ensemble Averaging of Transfer Learning Models for Identification of Nutritional Deficiency in Rice Plant." Accessed: Jul. 25, 2024. [Online]. Available: https://www.mdpi.com/2079-9292/11/1/148
[16] A. M. Rifai, S. Raharjo, E. Utami, and D. Ariatmanto, "Analysis for diagnosis of pneumonia symptoms using chest X-ray based on MobileNetV2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (CLAHE)," *Biomed. Signal Process. Control*, vol. 90, p. 105857, Apr. 2024, doi: 10.1016/j.bspc.2023.105857.
[17] M. Hayati *et al.*, "Impact of CLAHE-based image enhancement for diabetic retinopathy classification through deep learning," *Procedia Comput. Sci.*, vol. 216, pp. 57–66, Jan. 2023, doi: 10.1016/j.procs.2022.12.111.
[18] B. A. R. Hassan and F. A. A. Dawood, "Face-based Gender Classification Using Deep Learning Model," *J. Eng.*, vol. 30, no. 01, pp. 106–123, Jan. 2024, doi: 10.31026/j.eng.2024.01.07.
[19] S. Saifullah, A. P. Suryotomo, R. Dreżewski, R. Tanone, and T. Tundo, "Optimizing Brain Tumor Segmentation Through CNN U-Net with CLAHE-HE Image Enhancement," presented at the 2023 1st International Conference on Advanced Informatics and Intelligent Information Systems (ICAI3S 2023), Atlantis Press, Feb. 2024, pp. 90–101. doi: 10.2991/978-94-6463-366-5_9.
[20] "Brain Sciences | Free Full-Text | Brain Tumor Classification from MRI Using Image Enhancement and Convolutional Neural Network Techniques." Accessed: Jul. 25, 2024. [Online]. Available: https://www.mdpi.com/2076-3425/13/9/1320
[21] P. Macsik, J. Pavlovicova, S. Kajan, J. Goga, and V. Kurilova, "Image preprocessing-based ensemble deep learning classification of diabetic retinopathy," *IET Image Process.*, vol. 18, no. 3, pp. 807–828, 2024, doi: 10.1049/ipr2.12987.
[22] R. Sundar *et al.*, "Enhancing breast cancer detection from histopathology images: A novel ensemble approach with deep learning-based feature extraction," *MATEC Web Conf.*, vol. 392, p. 01139, 2024, doi: 10.1051/matecconf/202439201139.
[23] Z. Mahmood, K. Khan, M. Shahzad, A. Fayyaz, and U. Khan, "Enhanced detection and recognition system for vehicles and drivers using multi-scale retinex guided filter and machine learning," *Multimed. Tools Appl.*, vol. 83, no. 6, pp. 15785–15824, Feb. 2024, doi: 10.1007/s11042-023-16140-z.
[24] "CbcErDL: Classification of breast cancer from mammograms using enhance image reduction and deep learning framework | Multimedia Tools and Applications." Accessed: Jul. 25, 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11042-024-19616-8
[25] R. Prince, Z. Niu, Z. Y. Khan, M. Emmanuel, and N. Patrick, "COVID-19 detection from chest X-ray images using CLAHE-YCrCb, LBP, and machine learning algorithms," *BMC Bioinformatics*, vol. 25, no. 1, p. 28, Jan. 2024, doi: 10.1186/s12859-023-05427-5.
[26] M. Patel *et al.*, "Explainable AI for gastrointestinal disease diagnosis in telesurgery Healthcare 4.0," *Comput. Electr. Eng.*, vol. 118, p. 109414, Sep. 2024, doi: 10.1016/j.compeleceng.2024.109414.
[27] S. Nikitha, S. Prabhanjan, T. R. Rupa, and R. Dinesh, "Enhancing plant nutritional deficiency analysis: a multi-attention convolutional neural network approach," Multimed. Tools Appl., Sep. 2024, doi: 10.1007/s11042-024-20233-8.
[28] P. Singh and B. Dadhe, "Essential Mineral Nutrients for Plant Growth: Nutrient Functions and Deficiency Symptoms," 2022.
[29] "Nutrient-Deficiency-Symptoms-in-Rice." Accessed: Jul. 25, 2024. [Online]. Available: https://www.kaggle.com/datasets/guy007/nutrientdeficiencysymptomsinrice

[30] R. C. Gonzalez, R. E. Woods, and B. R. Masters, "Digital Image Processing, Third Edition," *J. Biomed. Opt.*, vol. 14, no. 2, p. 029901, 2009, doi: 10.1117/1.3115362.

[31] P. S. Heckbert, *Graphics gems IV*. in The Graphics gems series. Boston: AP Professional, 1994.

[32] E. H. Land, "The Retinex Theory of f Vision," *Sci. Am.*, vol. 237, no. 6, pp. 108–128, Dec. 1977, doi: 10.1038/scientificamerican1277-108.

[33] "Densely Connected Convolutional Networks | IEEE Conference Publication | IEEE Xplore." Accessed: Jul. 25, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8099726

[34] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 16, 2017, *arXiv*: arXiv:1704.04861. doi: 10.48550/arXiv.1704.04861.

[35] A. Vaswani *et al.*, "Attention Is All You Need," Aug. 01, 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," vol. 8691, 2014, pp. 346–361. doi: 10.1007/978-3-319-10578-9_23.