DELG Net: A Dual-Stream Cross-Attention Framework for Automated Nutrient Deficiency Detection in Mulberry Leaves

S. Raghavendrachar
Department of Computer Science and Engineering, K. S. Institute of Technology, Karnataka – 560109, India, Affiliated to VTU, Belagavi, Visvesveraya Technological University, Belagavi – 590018, India
raghavendrachars@ksit.edu.in (corresponding author)

Rekha B. Venkatapur
Department of Computer Science and Engineering, K. S. Institute of Technology, Karnataka – 560109, India, Affiliated to VTU, Belagavi, Visvesveraya Technological University, Belagavi – 590018, India
rekhabvenkatapur@ksit.edu.in

V. Karthik
Department of Computer Science and Engineering, K. S. Institute of Technology, Karnataka – 560109, India, Affiliated to VTU, Belagavi, Visvesveraya Technological University, Belagavi – 590018, India
Karthik.venkatu@gmail.com

Received: 13 January 2026 | Revised: 2 February 2026 | Accepted: 14 February 2026

ABSTRACT

Nutrient deficiencies in mulberry leaves play a decisive role in sericulture, since they directly affect the growth of silkworms and the production of silk. Traditional visual observation by professionals is subjective, laborious, and cannot be used in massive monitoring. Although generic Convolutional Neural Network (CNN) models are frequently applied to this task, they often struggle to capture the fine-grained spatial and textural details essential for identifying specific nutrient stress. To address these drawbacks, this paper presents DELG Net, a biologically-related two-stream convolutional model that conjoins global and regional feature processing. Macroscopic color and contour features are captured by the global stream, and vein and texture patterns are highlighted with high-pass filtering by the local stream. A cross-attention fusion model dynamically increases the contribution of the two streams according to the prominence of the symptoms, and feature weighting can be adjusted adaptively and contextually. Experiments on a curated dataset of mulberry plant leaves show that DELG Net achieves better precision and interpretability than state-of-the-art CNNs, with a total classification accuracy of 95.4% and consistent results in all three classes of nitrogen, potassium, and phosphorus deficiencies. The proposed model can be used to monitor nutrient stress in real-time and on a large scale to increase the efficiency of silk production.

Keywords-deep learning; cross-attention fusion; dual-stream architecture; plant health monitoring; precision agriculture

I. INTRODUCTION

Sericulture, which involves the cultivation of Bombyx mori, is a crucial agro-industry and rural economic sector that has a huge influence on the world textile market. Since silkworms only have mulberry leaves as food, the quality of the food directly influences the number of cocoons and the quality of their fibers. The lack of the necessary macronutrients, namely, Nitrogen (N), Phosphorus (P), and Potassium (K), hampers the growth of larvae and deteriorates the quality of silk, leading to significant economic loss [1]. Nitrogen is essential for protein and amino acid synthesis [2], whereas potassium controls enzyme activity, osmotic balance, and stress resistance [3].

Traditional diagnostic methods for nutritional deficiency are usually subjective, slow, and unreliable in manual visual examination. Although specialists can recognize signs such as nitrogen-induced chlorosis or potassium-induced marginal scorching, these visual clues tend to be unreliable in changing environmental conditions [4]. The industry is moving to Artificial Intelligence (AI) and Deep Learning (DL) to obtain objective and scalable diagnostics [5]. Convolutional Neural Networks (CNNs) such as ResNet, VGGNet, and MobileNet have performed well in general disease detection by extracting features from raw images [6, 7]. Nevertheless, such general architectures tend to be biologically uninterpretable and find it difficult to differentiate between subtle overlapping symptoms, including vein deformation, marginal curling, or certain color gradients, which are symptoms of different nutrient deficiencies, reducing accuracy in complex cases [8, 9].

CNDD-Net [10] is a ResNet-based CNN with CBAM (channel and spatial) attention to detect corn leaf disease and nutrient deficiency. Although this model is very accurate and less complex, it is not explicit in global-local contrastive feature modeling and continuous health score regression. In [11], an explainable CNN-Vision Transformer (ViT) architecture was proposed for the diagnosis of mulberry leaf disease, including Grad-CAM interpretability, obtaining 95.6% accuracy on the three disease categories. However, this method is disease-centric and fails to classify nutrient deficiency and include multi-scale feature fusion. In [12], the use of color and texture features as a form of nutrient deficiency detection was investigated with classical machine learning algorithms, including SVM, KNN, and Random Forest, although these hand-crafted feature-based algorithms did not demonstrate strong generalization to real-field environments. In [13], cotton leaf disease detection was based on the YOLOv5-based architecture, performing image preprocessing, and showing high performance compared to ResNet50 and VGG16 in terms of classification metrics and ROC analysis.

A hybrid-DAE-LSTM model [14] combined spatial and temporal data to diagnose nutrient status in aquaponic lettuce to achieve fine-grained nutrient differentiation; however, this strategy did not use curriculum learning to train the model progressively. A modified version of the YOLOv4 architecture, extended with dense DenseNet-based backbones, altered residual blocks, and improved feature fusion (SPP+PANet), was more effective in tasks with local texture and feature representation, but cannot be effectively considered when the task involves high-resolution fine-localization classification [15]. In [16], an unsupervised domain adaptation algorithm was based on matching source and target feature distributions to address environmental variability in crop-weed recognition, but without explicitly providing the categorization of cross-scale or hierarchical interactions between symptom characteristics. In [17], it was noted that deep and aggressively adaptive models can be constrained in their deployment in real-world scenarios due to computational, power, and latency constraints.

In [18], DL methods were explored to predict plant stress, including model architecture, fusion mechanism, domain adaptation, and deployment issues. An experimental investigation of a domain-specific non-invasive diagnosis of nutrient deficiency used CNNs on RGB crop images and recognized the challenge of low inter-class separability of deficiency patterns [19]. In [20], four DCNNs, namely Inception-v3, ResNet50, NASNet, and DenseNet121, were compared in detecting ten rice leaf nutrient deficiencies with more than 90% test accuracy, proving the usefulness of deep models compared to classical methods. Multi-modal fusion showed the advantage of the DL-CRoP framework in tomato and maize to identify nitrogen deficiency using RGB and root imaging with a high accuracy (approximately 93%) [21]. However, despite these developments, the incorporation of biologically compatible, interpretable global-local feature cues is still a standing problem.

Multi-scale DL is an effective tool in improving subtle pattern discovery, using features at varying scales [22]. MFFTNet [23], a multi-scale fusion and Transformer-based framework, extracts hierarchical features from remote-sensing images that fit well with leaf-level stress analysis. In [24], a few-shot plant stress classification model used cascaded multi-scale features and channel attention to facilitate generalization when using a limited amount of data. In crop image semantic segmentation, multi-scale feature fusion was used in [25] to enhance the results to delineate healthy and diseased areas. In [26], multimodal fusion approaches were used to monitor plants based on RGB, multispectral, thermal, and soil information, with an emphasis on practical implementation issues. In [27], smart agriculture systems were reviewed, highlighting multimodal sensing to differentiate between nutrient stress, drought, and disease. In [28], a hybrid between manifold learning and intermediate fusion allowed controlled latent-space multimodal integration, which can be used to combine image streams with auxiliary data.

Most CNN-based models for nutrient deficiency detection operate in an image-based manner, without processing images into explicit global and local components (color and morphology, and venation and texture, respectively), leading to an unbiologically sensitive representation of features and sensitivity to small-scale features of nutrient stress. Although multi-scale feature-fusion has been employed to improve coarse-to-fine feature learning, less explored is cross-attention-based fusion, designed to support nutrient deficiency analysis. Simultaneously, as the trend to focus more on explainable agricultural AI develops, there is an opportunity to explore more interpretable models by displaying visual reasoning, which might be equally desirable with high accuracy. In order to fill these gaps, this paper proposes a self-organizing DELG Net, which is a dual-stream bio-inspired architecture that expresses both global and local features and warps them together using cross-attention fusion. This design is lightweight, with accuracy, interpretability, and computation efficiency allowing for real-time on-field deployment.

II. PROPOSED METHODOLOGY

DELG Net is a DL model for identifying and measuring nutrient deficiencies in mulberry leaves (Morus spp.). The architecture simulates agronomic visual inspection through joint intensity modeling of global leaf properties (color and shape) and the local structural clues (venation and texture) in two parallel processing streams. To facilitate a coherent global-local interaction of features, these streams are combined with the help of a cross-attention fusion mechanism. DELG Net uses the dual-head training approach, performing both nutrient deficiency classification and continuous leaf health score regression. The designed pipeline, shown in Figure 1, is able to learn both macro- and micro-level elements concurrently.

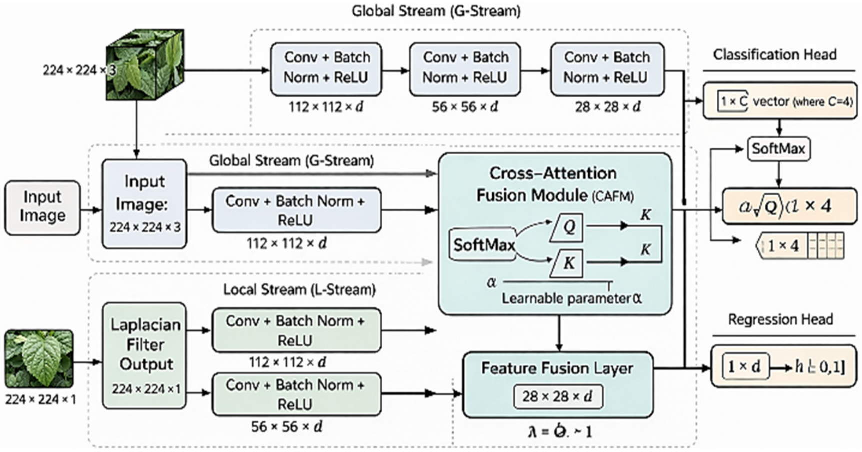

Fig. 1. Workflow of the DELG-Net model

A. Dataset and Preprocessing

Images of mulberry leaves were divided into four nutrient conditions, namely, Healthy, Nitrogen-deficient (N), Phosphorus-deficient (P), and Potassium-deficient (K). Photos were taken from a silkworm house in Keralalusandra, Karnataka, India, with natural field-level variability of leaf morphology and pigmentation. Images were captured at daylight, using a Sony Cybershot digital camera with a 20.1 MP sensor. The filtered dataset has 4,600 images, allotted in Healthy (1,000), N-deficient (1,200), P-deficient (1,200), and K-deficient (1,200). All classes have characteristic phenotypic features, such as dark green coloration in all healthy leaves, class-specific chlorosis, pigment-abnormalities and marginal necrosis in nutrient-depleted samples, which give discriminatory visual cues to learning-based classification. To enable continuous health score regression, a ground truth health label $h \in [0,1]$ was assigned to each image in the training set. Three agronomic experts assigned labels manually according to a standardized visual scoring rubric. Healthy leaves with complete dark green pigmentation were awarded a score of 1.0. A scale of deficient leaves was based on the percentage of the area of the symptomatic leaf (chlorosis, necrosis, or marginal curling): early-stage symptoms received a score of between 0.6 and 0.8, and severe symptoms were assessed with a score of between 0.1 and 0.3. The label of every image was the average score of the experts to provide objective regression training. The dataset was stratified into training (70%), validation (15%), and test (15%) subsets to maintain class balance. Data augmentation strategies, such as random flipping, rotation, and scaling, were used to represent real-world variability, avoid overfitting, and improve generalization. The images were all converted to RGB, resized to 224×224, and normalized using the mean value. This uniform preprocessing not only provided the model with stable training but also with faster convergence, emphasizing its stability under different visual conditions. The filtered image I_L is calculated using convolution, $I_L = I * K$, where K is the Laplacian kernel matrix and $*$ is the convolution operator. A 3×3 Laplacian kernel with a center coefficient of -4 was used to effectively extract high-frequency components while maintaining computational efficiency.

B. Dual-Stream Feature Extraction

DELG-Net uses two parallel convolutional streams, a Global Stream (G-Stream) and a Local Stream (L-Stream), to cooperatively acquire holistic and finer leaf features. The G-Stream manipulates the primary RGB image to capture global contextual information, such as color distribution, general shape, and boundary structure. These visualizations allow useful detection of symptoms of color-dominant nutrient deficiencies, such as chlorosis due to nitrogen stress, since they model the large-scale spatial patterns that indicate the overall leaf health. Conversely, the L-Stream acts on the Laplacian-filtered image I_L that highlights all details due to high frequency, such as the structures of veins and variation in texture at edges. This stream is especially created to identify subtle local manifestations, such as vein deformity, marginal curling, and edge necrosis, which are normally linked to a vitamin deficit, such as potassium deficiency. Both streams consist of repeated blocks of convolution-batch normalization-ReLU, and are then followed by max pooling blocks that successively decrease the size of the space and maintain the important features. The two streams provide the global feature map (F_g) and the local feature map (F_l), which are combined to give a complete view of the health status of the leaf, both in the macroscopic image and the microscopic texture.

$$F_g = CNN_g(I), \quad F_l = CNN_l(I_L) \tag{1}$$

C. Cross-Attention Fusion Mechanism

A Cross-Attention Fusion Module (CAFM) incorporates the two complementary aspects of the global and local streams by adapting and highlighting the most salient symptom-specific representations. The module maps embedding of query features into a Query-Key-Value (QKV) space, which allows adaptive interaction of features by attention according to visual relevance:

$$Q_g = W_Q F_Q, \; K_l = W_K F_l \tag{2}$$

where W_Q and W_K are learnable projection matrices. The attention weights A are calculated using a scaled dot-product attention:

$$A = SoftMax\left(\frac{Q_g K_l^T}{\sqrt{d}}\right) \tag{3}$$

where d is the dimensional scaling factor. The fused feature map F_f is computed as a weighted combination of global and local streams:

$$F_f = \alpha F_g + (1-\alpha)(AF_l) \tag{4}$$

Here, α is a learnable parameter that dynamically balances the global-local feature contributions. This ensures that the model focuses on color-based cues for nitrogen deficiency and vein-based cues for potassium deficiency as required.

D. *Feature Fusion and Output Heads*

The resulting feature vectors F_f from the dual streams are then fused and fed through fully connected layers, resulting in a composite representation of the visual features of the leaf. Based on such a common representation, two output branches are obtained, one classified and the other regressed. The Classification Head uses a SoftMax layer to estimate the nutrient deficiency category out of C predefined categories, i.e., Healthy, Nitrogen-deficient, Phosphorus-deficient, and Potassium-deficient. Loss associated with classification is computed using the categorical cross-entropy function:

$$L_{cls} = \sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{5}$$

where y_i represents the true class label and $\hat{y}_i$ is the predicted probability for each class.

The Regression Head uses a sigmoid activation function to output a continuous health score $h \in [0,1]$, where values close to 1.0 indicate a healthy leaf and values near 0.0 correspond to severe deficiency. The regression loss is measured using Mean Squared Error (MSE):

$$L_{reg} = \frac{1}{N}\sum_{i=1}^{N}(h_i - \hat{h}_i)^2 \tag{6}$$

Finally, the combined loss function integrates both objectives, balancing classification accuracy and health estimation through weighted coefficients λ_1 and λ_2:

$$L_{Total} = \lambda_1 L_{cls} + \lambda_2 L_{reg} \tag{7}$$

This joint optimization strategy enables the model to simultaneously identify the type of nutrient deficiency and provide a continuous measure of leaf health, improving both interpretability and robustness.

E. *Training Strategy*

The DELG-Net model was trained using the Adam optimizer to ensure stable convergence and adaptive learning. The training process utilized an initial learning rate of 1×10^{-4}, a batch size of 32, and a total of 100 epochs. The optimizer parameters were set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Early stopping was employed based on validation loss with a patience of 10 epochs to avoid overfitting. Furthermore, dropout layers with a probability of 0.3 were introduced in the fully connected stages to enhance generalization. The implementation was carried out in TensorFlow 2.12 and PyTorch 2.1 environments, and the training was performed on a workstation powered by an NVIDIA RTX 3080 GPU. Table I summarizes the training environment.

TABLE I. TRAINING ENVIRONMENT PARAMETERS

Parameter	Description / Value
Optimizer	Adam
Initial Learning Rate	1×10^{-4}
Batch Size	32
Number of Epochs	100
Optimizer Parameters	$\beta_1 = 0.9$ and $\beta_2 = 0.999$
Early Stopping	Based on validation loss (patience = 10 epochs)
Dropout Rate (p)	0.3 (in fully connected layers)
Framework	TensorFlow 2.12 / PyTorch 2.1
Hardware Used	NVIDIA RTX 3080 GPU

III. RESULTS AND DISCUSSION

A. *Quantitative Results*

The performance evaluation results show that DELG-Net can classify and differentiate types of nutrient deficiency in the mulberry leaf, with a high generalizability among the four classes, as shown in Table II. The highest false-positive rates are observed in the Healthy and Nitrogen-deficient classes because they have different visual features. Minor changes in the precision of Phosphorus deficiency are caused by features that overlap with Potassium deficiency, but the overall F1-score of 0.907 represents a good and balanced overall predictive ability over the nutrient groups.

TABLE II. CLASS-WISE PERFORMANCE OF DELG NET

Class	Precision	Recall	F1-score	Accuracy
Healthy	**0.962**	**0.93**	**0.924**	0.953
Nitrogen-Deficient (N)	0.93	0.922	0.924	0.953
Phosphorus-Deficient (P)	0.89	0.91	0.907	0.953
Potassium-Deficient (K)	0.94	0.9	0.92	0.953

The overall classification accuracy of 95.3% indicates a high level of discrimination under different visual conditions. In addition, the macro-averaged precision (0.93), recall (0.915), and F1-score (0.918) values show that there is a balanced tradeoff between sensitivity and specificity, and the false positives and missed detections are minimized. The findings support the usefulness of the suggested dual-stream global-local feature learning approach to improve the performance and interpretability of nutrient deficiency patterns in leaf pictures.

DELG Net was evaluated against MobileNetV2 and YOLOv8 under identical preprocessing and training settings. As summarized in Table III, DELG-Net outperformed both baselines across all key metrics, achieving an accuracy of 95.3%, exceeding MobileNetV2 and YOLOv8 by 9.4% and 5%, respectively. This improvement stems from its dual-stream architecture, which jointly models global contextual features (color distribution and leaf morphology) and local structural cues (venation and marginal stress patterns). In contrast, MobileNetV2's single-stream lightweight design limits fine-grained feature learning, resulting in lower recall and F1-scores, while YOLOv8's object-detection-oriented optimization prioritizes bounding-box localization and speed over pixel-level nutrient discrimination. The precision, recall, and F1-score of DELG Net are high (0.93, 0.915, and 0.918), indicating balanced and reliable discrimination of the healthy and nutrient-deficient samples in all classes.

TABLE III. PERFORMANCE COMPARISON OF DELG NET WITH BASELINE MODELS

Model	Class	Precision	Recall	F1-score	Accuracy
DELG Net (Proposed)	Healthy	0.962	0.93	0.924	0.953
	Nitrogen	0.93	0.922	0.924	
	Potassium	0.94	0.90	0.92	
	Phosphorus	0.89	0.91	0.907	
MobileNet	Healthy	0.89	0.87	0.876	0.859
	Nitrogen	0.83	0.80	0.82	
	Potassium	0.84	0.832	0.83	
	Phosphorus	0.78	0.76	0.774	
YOLOv8	Healthy	0.943	0.91	0.93	0.904
	Nitrogen	0.89	0.87	0.875	
	Potassium	0.87	0.84	0.86	
	Phosphorus	0.79	0.78	0.79	

These findings affirm the usefulness of the proposed cross-attention fusion mechanism in modeling global and local features simultaneously, resulting in higher discriminability and interpretability. In general, DELG-Net performs better than standard CNNs and object-detection models with regard to the ability to reproduce subtle and overlapping patterns of nutrient deficiency. The dual-stream design generalizes across leaf thickness and serration of mulberry leaf varieties despite differences in structural variability, since the global color cues (deficit signals, with them) are encoded in the local venation texture, although the current dataset was gathered in control field conditions.

B. Qualitative Analysis

Figure 2 shows a Grad-CAM visualization, demonstrating that the proposed DELG Net can be understood in biological terms and can highlight biologically significant areas, namely venation, leaf margins, and discoloration spots. In healthy leaves, all attention is evenly distributed, but in samples with nutrient deficiencies, activations are strong in the chlorotic and necrotic areas, which characterizes the good localization of symptoms. This correspondence between model attention and agronomic symptom patterns justifies the usefulness of the cross-attention fusion mechanism in the context of linking computational learning with biological significance. Figure 2(a) presents the initial RGB image of a healthy mulberry leaf with even green color and smooth surface, which are signs of proper nutrition. The Laplacian-improved image in Figure 2(b) highlights high-frequency information, veins, and edges, helping to realize the early existence of small structural anomalies. Figure 2(c) is the result of Grad-CAM, with warmer areas indicating the greater importance of features. Its specific and balanced central lamina activation reaffirms that DELG Net can be trained on spatially consistent and diagnostically significant features without classification bias.

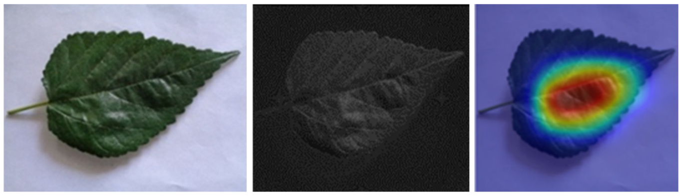
(a) (b) (c) Grad-CAM

Fig. 2. (a) Health leaf, (b) Laplacian filtered, (c) Grad-CAM maps.

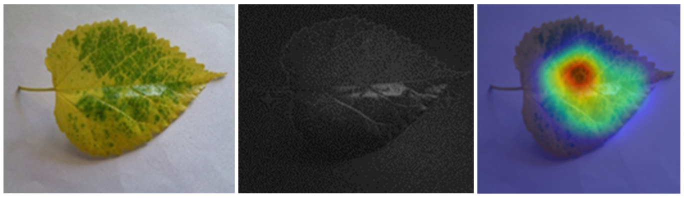
(a) (b) (c)

Fig. 3. Model's response to a Nitrogen-deficient leaf: (a) Nitrogen-deficient leaf, (b) Laplacian-filtered, (c) Grad-CAM filtered.

Figure 3 shows the reaction of DELG Net to nitrogen-deficient mulberry leaves, demonstrating that it is capable of capturing biologically relevant stress patterns. Nitrogen deficiencies appear as uniform chlorosis moving outward from the older regions to the central lamina, as shown in Figure 3(a), a secondary diagnostic symptom. Figure 3(b) shows the Laplacian-emphasized image, which enhances high-frequency information, including the shape of veins and edge outlines, and helps identify the slightest stress-related changes in the texture at the level of the color difference in RGB. The underlying Grad-CAM map in Figure 3(c) shows focused activations on chlorotic areas of the central lamina, with the warmer colors being the ones that are more relevant. Such a localized attention proves that the cross-attention fusion mechanism is successful in linking nutrient stress to color and texture information, allowing for biologically appropriate model decisions.

The visualization can affirm that DELG Net works with adaptive analysis and pays attention to agriculturally relevant features, including vein discoloration and interveinal chlorosis, rather than using global image cues. The high level of correspondence between biological symptoms and the activation regions is an indication of the interpretability and diagnostic reliability of the model when it comes to the detection of nitrogen deficiency. The confusion matrix in Figure 4 also indicates the strong classification ability for all four classes, as indicated by the high degree of diagonal dominance and small misclassification. The most accurate were healthy leaves (96%), with future pigment and texture patterns being similar. Potassium-deficient samples were classified with 95% accuracy, showing that marginal necrosis and vein-edge variants were well captured. Leaves with Nitrogen-deficiency were classified with an accuracy of 93%, with a slight degree of confusion with the Healthy category, whereas Phosphorus-deficient samples had 91% accuracy, with a slight level of confusion with Potassium-deficiency because of the similarity in patterns of interveinal discoloration.

The confusion matrix analysis establishes that DELG Net has balanced discriminative capabilities in each category of nutrient, with a small number of off-diagonal errors, demonstrating success in distinguishing patterns of deficiency with similar appearances using its two-stream structure. These findings indicate the strength, stability, and dependability of the model on the diagnosis of nutrient stress of plant leaves in the real world. In comparison, the MobileNetV2 confusion matrix (Figure 5) indicates average performance, especially on overlapping deficiency symptoms because of the small fine-grained feature representation. Accuracies of 88%, 81%, 84%, and 74% for healthy, nitrogen, potassium, and phosphorus deficiencies, respectively, indicated that MobileNetV2 was less sensitive to minor spectral and structural stress effects.

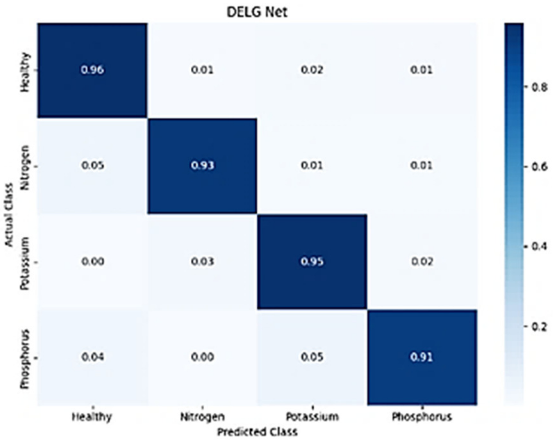

Fig. 4. DELG-Net confusion matrix.

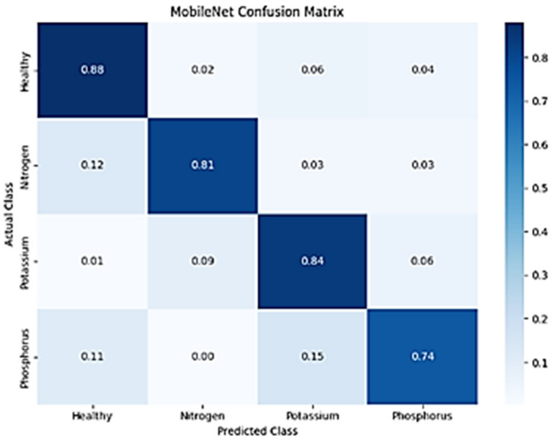

Fig. 5. MobileNetV2 confusion matrix.

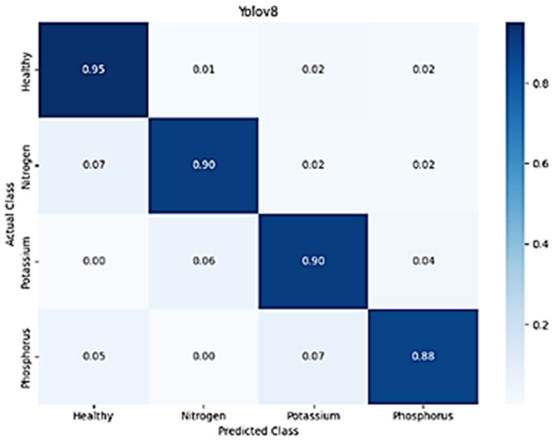

Fig. 6. YOLOv8 confusion matrix.

The results indicate that although MobileNetV2 is computationally efficient, its single-stream design limits the capture of multi-scale context and fine texture details. A stronger local feature extraction mechanism is needed to distinguish nutrient classes with similar symptoms. Hence, advanced dual-stream models such as DELG Net, which fuse global color and local morphology, offer superior precision and interpretability.

Figure 6 shows that YOLOv8 provides competitive performance bus has drawbacks in terms of fine-grained classification of nutrient deficiency because it is an object-detection-based architecture. Although it is very precise with healthy leaves (95%), it is sensitive to finer changes in pixels, which limits its sensitivity to nitrogen (90%), potassium (90%), and phosphorus (88%) deficiencies, with visible inter-class confusion. YOLOv8 depends on bounding-box-level features instead of more detailed venation and edge textures that are needed to distinguish between different nutrients. YOLOv8 is not as effective as DELG-Net, which has a dual-stream framework and cross-attention fusion that allow superior global and local features capture. In general, the findings back the claim that task-specific architectures such as DELG-Net are more appropriate to detect subtle variation patterns in nutrient variations as compared to generic object detectors.

Figure 7 contrasts DELG-Net, MobileNetV2, and YOLOv8 in four nutrient conditions. DELG-Net is significantly ahead of the others as it reaches a consensus of more than 0.9 in all classes. Its two-stream combination of RGB-based global color and Laplacian-filtered local texture can reasonably predict finer symptoms, including marginal necrosis, interveinal chlorosis, and gradients of discoloration changes that single-stream models are mostly not able to detect.

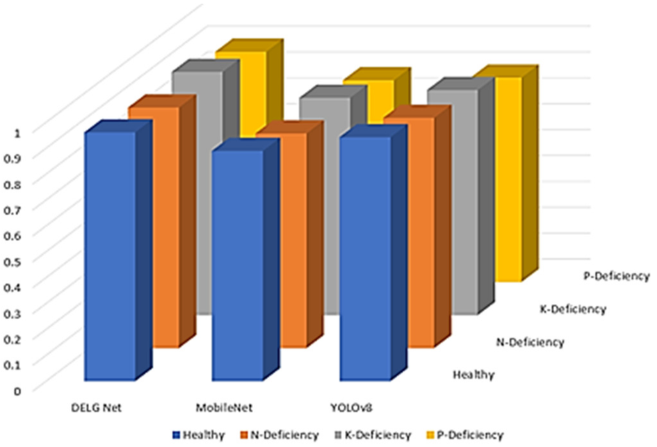

Fig. 7. Class-wise precision comparison.

MobileNetV2 experiences a problem with phosphorus and potassium deficiencies due to its small receptive areas, and YOLOv8 is oriented towards object localization. The cross-attention and two-stream approach of DELG-Net allows the assessment of nutrients in the environment robustly, interpretably, and effectively.

IV. CONCLUSION

This study introduced DELG Net, a biologically inspired dual stream deep learning architecture, to perform automated detection and quantification of nutrient deficiencies in Morus leaves. DELG Net has better interpretability and diagnostic reliability by combining the global color and local texture reasoning with a cross-Attention fusion mechanism. Its accuracy of 95.4% on experiments on 4,600 leaf images surpassed MobileNetV2 and YOLOv8. Its two-output architecture, which integrates categorical classification and continuous health scoring, can be used to assist in precision sericulture and monitor numerous plants. Grad-CAM revealed that agronomically relevant areas were paid attention to. In addition to technical benefits, DELG-Net facilitates sustainable agriculture by providing opportunities to track nutrient-induced stress in crops early, effectively utilize fertilizers, and combine them with real-time mobile or edge-AI applications.

REFERENCES

[1] E. A. Aldakheel, M. Zakariah, and A. H. Alabdalall, "Detection and identification of plant leaf diseases using YOLOv4," *Frontiers in Plant Science*, vol. 15, Apr. 2024, Art. no. 1355941, https://doi.org/10.3389/fpls.2024.1355941.

[2] K. Venkatesh and K. J. Naik, "Nutrient deficiency identification and yield-loss prediction in leaf images of groundnut crop using transfer learning," *Signal, Image and Video Processing*, vol. 18, no. 5, pp. 4553–4568, July 2024, https://doi.org/10.1007/s11760-024-03094-4.

[3] M. H. Bijoy *et al.*, "Towards Sustainable Agriculture: A Novel Approach for Rice Leaf Disease Detection Using dCNN and Enhanced Dataset," *IEEE Access*, vol. 12, pp. 34174–34191, 2024, https://doi.org/10.1109/ACCESS.2024.3371511.

[4] J. Fuentes-Pacheco *et al.*, "A Curriculum Learning Approach to Classify Nitrogen Concentration in Greenhouse Basil Plants Using a Very Small Dataset and Low-Cost RGB Images," *IEEE Access*, vol. 12, pp. 27411–27425, 2024, https://doi.org/10.1109/ACCESS.2024.3367614.

[5] M. Majdalawieh, S. Khan, and T. Islam, "Using Deep Learning Model to Identify Iron Chlorosis in Plants," *IEEE Access*, vol. 11, pp. 46949–46955, 2023, https://doi.org/10.1109/ACCESS.2023.3273607.

[6] M. Shoaib *et al.*, "An advanced deep learning models-based plant disease detection: A review of recent research," *Frontiers in Plant Science*, vol. 14, Mar. 2023, Art. no. 1158933, https://doi.org/10.3389/fpls.2023.1158933.

[7] A. Bera, D. Bhattacharjee, and O. Krejcar, "PND-Net: plant nutrition deficiency and disease classification using graph convolutional network," *Scientific Reports*, vol. 14, no. 1, July 2024, Art. no. 15537, https://doi.org/10.1038/s41598-024-66543-7.

[8] K. Wei *et al.*, "Explainable Deep Learning Study for Leaf Disease Classification," *Agronomy*, vol. 12, no. 5, Apr. 2022, Art. no. 1035, https://doi.org/10.3390/agronomy12051035.

[9] H. K. Kondaveeti and C. G. Simhadri, "Evaluation of deep learning models using explainable AI with qualitative and quantitative analysis for rice leaf disease detection," *Scientific Reports*, vol. 15, no. 1, Aug. 2025, Art. no. 31850, https://doi.org/10.1038/s41598-025-14306-3.

[10] S. Timilsina *et al.*, "CNDD-Net: A Lightweight Attention-Based Convolutional Neural Network for Classifying Corn Nutritional Deficiencies and Leaf Diseases," *Electronics*, vol. 14, no. 7, Apr. 2025, Art. no. 1482, https://doi.org/10.3390/electronics14071482.

[11] M. A. Hasan *et al.*, "Mulberry leaf disease detection by CNN-ViT with XAI integration," *PLOS One*, vol. 20, no. 6, June 2025, Art. no. e0325188, https://doi.org/10.1371/journal.pone.0325188.

[12] J. G. A. Barbedo, "Detection of nutrition deficiencies in plants using proximal images and machine learning: A review," *Computers and Electronics in Agriculture*, vol. 162, pp. 482–492, July 2019, https://doi.org/10.1016/j.compag.2019.04.035.

[13] S. Madhu and V. RaviSankar, "Comprehensive Analysis of a YOLO-based Deep Learning Model for Cotton Plant Leaf Disease Detection," *Engineering, Technology & Applied Science Research*, vol. 15, no. 1, pp. 19947–19952, Feb. 2025, https://doi.org/10.48084/etasr.8944.

[14] M. F. Taha *et al.*, "Deep Learning-Enabled Dynamic Model for Nutrient Status Detection of Aquaponically Grown Plants," *Agronomy*, vol. 14, no. 10, Oct. 2024, Art. no. 2290, https://doi.org/10.3390/agronomy14102290.

[15] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network." arXiv, 2021, https://doi.org/10.48550/ARXIV.2111.00298.

[16] T. Ilyas, J. Lee, O. Won, Y. Jeong, and H. Kim, "Overcoming field variability: unsupervised domain adaptation for enhanced crop-weed recognition in diverse farmlands," *Frontiers in Plant Science*, vol. 14, Aug. 2023, Art. no. 1234616, https://doi.org/10.3389/fpls.2023.1234616.

[17] G. Ignesti, D. Moroni, and M. Martinelli, "Towards the actual deployment of robust, adaptable, and maintainable AI models for sustainable agriculture," presented at the 19th Conference on Computer Science and Intelligence Systems, Belgrade, Serbia, Nov. 2024, pp. 33–39, https://doi.org/10.15439/2024F2991.

[18] N. Paul, G. C. Sunil, D. Horvath, and X. Sun, "Deep learning for plant stress detection: A comprehensive review of technologies, challenges, and future directions," *Computers and Electronics in Agriculture*, vol. 229, Feb. 2025, Art. no. 109734, https://doi.org/10.1016/j.compag.2024.109734.

[19] J. Yi *et al.*, "Deep Learning for Non-Invasive Diagnosis of Nutrient Deficiencies in Sugar Beet Using RGB Images," *Sensors*, vol. 20, no. 20, Oct. 2020, Art. no. 5893, https://doi.org/10.3390/s20205893.

[20] Z. Xu *et al.*, "Using Deep Convolutional Neural Networks for Image-Based Diagnosis of Nutrient Deficiencies in Rice," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–12, Aug. 2020, https://doi.org/10.1155/2020/7307252.

[21] M. Urfan *et al.*, "The Deep Learning-Crop Platform (DL-CRoP): For Species-Level Identification and Nutrient Status of Agricultural Crops," *Research*, vol. 7, Jan. 2024, Art. no. 0491, https://doi.org/10.34133/research.0491.

[22] E. Elizar, M. A. Zulkifley, R. Muharar, M. H. M. Zaman, and S. M. Mustaza, "A Review on Multiscale-Deep-Learning Applications," *Sensors*, vol. 22, no. 19, Sept. 2022, Art. no. 7384, https://doi.org/10.3390/s22197384.

[23] Y. Cheng *et al.*, "Multi-scale Feature Fusion and Transformer Network for urban green space segmentation from high-resolution remote sensing images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 124, Nov. 2023, Art. no. 103514, https://doi.org/10.1016/j.jag.2023.103514.

[24] H. Lin, R. Tse, S. K. Tang, Z. Qiang, and G. Pau, "Few-shot learning approach with multi-scale feature fusion and attention for plant disease recognition," *Frontiers in Plant Science*, vol. 13, Sept. 2022, Art. no. 907916, https://doi.org/10.3389/fpls.2022.907916.

[25] G. Zheng, Z. Jiang, X. Zhang, and D. Jiang, "Multi-scale feature fusion-based semantic segmentation network for agricultural remote sensing images," *Chemical and Biological Technologies in Agriculture*, vol. 12, no. 1, Aug. 2025, Art. no. 126, https://doi.org/10.1186/s40538-025-00833-8.

[26] Z. X. Yang, Y. Li, R. F. Wang, P. Hu, and W. H. Su, "Deep Learning in Multimodal Fusion for Sustainable Plant Care: A Comprehensive Review," *Sustainability*, vol. 17, no. 12, June 2025, Art. no. 5255, https://doi.org/10.3390/su17125255.

[27] T. Ali *et al.*, "Smart agriculture: utilizing machine learning and deep learning for drought stress identification in crops," *Scientific Reports*, vol. 14, no. 1, Dec. 2024, Art. no. 30062, https://doi.org/10.1038/s41598-024-74127-8.

[28] M. Bodaghi, M. Hosseini, and R. Gottumukkala, "A Multimodal Intermediate Fusion Network with Manifold Learning for Stress Detection." arXiv, 2024, https://doi.org/10.48550/ARXIV.2403.08077.