Rachit Khandelwala , 2026, 14:4
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Towards Sustainable Agriculture: A Hybrid Deep Learning with Super Resolution for Grape Leaf Disease Detection from Low Resolution Image

Rachit Khandelwala, Paresh Jain
Department of Electronics and Communication Engineering, Suresh Gyan Vihar University, Jaipur, India

Abstract- Grapes are small, sweet, and versatile fruits that come in various colors and are not only enjoyed fresh but also used for making wine, raisins, and a variety of culinary products. A healthy grape harvest requires prompt identification of grape disease and taking appropriate measures to stop it from spreading. Accurate disease identification is now possible with the help of recent research advancements. However, these techniques yield low accuracy and perform poorly with low-resolution images. A novel framework incorporating TSRNet model based super resolution followed by DataLiteViT classification model is suggested as a solution to this problem. For scale factors of ×2, ×4, and ×6, respectively, the utilised lightweight TSRNet model with 2.25 M parameters can reach 30.12, 28.38, and 27.57dB Peak Signal Noise Ratio (PSNR) values and Structural Similarity Index Measure (SSIM) value of 0.913, 0.841, and 0.756. The proposed DataLiteViT model achieved accuracies of 97.90%, 97.12%, and 96.69% for low resolution images with scale factors of × 2, × 4, and × 6, respectively. Furthermore, the employed method utilizing TSRNet model for super-resolution improved DataLiteViT's classification accuracies of 99%, 98.71%, and 98.28% for LR images with scale factors of ×2, ×4, and ×6, demonstrating an efficient and highly effective solution.

Keywords: Ai in agriculture, image super resolution, transformer, interpolation, grape leaf disease detection, accuracy parameters

I. INTRODUCTION

Indian Council for Agricultural Research (ICAR) states that India is the sixth largest producer of grapes worldwide. About 53,000 tonnes of grapes are exported from India annually, making up 1.5% of all grape exports worldwide [1]. As such, they are an important crop for expanding the nation's economy. The prompt identification of crop diseases is essential for a healthy harvest, limiting yield losses, and minimizing the necessity for pesticides[2]. Conventional disease diagnosis techniques depend on grower expertise or professional advice, which is a tedious and inefficient procedure.

Recent digitization of agricultural images [3] has significantly advanced semi-automatic [4], [5] and completely automatic [6], [7], [8] disease diagnosis in agriculture. Various deep learning models based on CNNs [9] can automatically extract key features like color, texture etc. to efficiently detect diseases. Recent research investigations show that ViT-based networks outperform CNN on categorization tasks. Unfortunately, they encounter challenges with capturing local dependencies on limited datasets often used in agricultural domain and require extensive datasets [10] such as JFT-300M [11] to achieve optimal performance. Various attempts have been made to overcome these limitations by merging ViTs with CNNs [12], [13], [14] and adapting the self-attention process [15], [16], [17], [18], [19] to cope with medium-sized datasets like ImageNet[20]. However, their performance on smaller dataset remains a challenge. The proposed methodology addresses this issue by improving receptive fields of DataLiteViT based on a novel Displaced Patch Embedding technique.

Furthermore, despite achieving remarkable classification accuracy on high-resolution input images, the performance of the various deep learning approaches significantly declined for small target detection [21], lower resolution images [22] and environmental imaging constraints [23]. Early research in smart farming using UAVs to identify plant diseases through images [24], [25] has shown that, although increasing altitude expands field coverage, it also makes it more challenging to identify different plant characteristics (such as lesions, etc.). Furthermore, drone propeller motion, climatic conditions, and low light levels can all contribute to image deterioration, including low resolution [26], blurring, etc. In order to resolve such issues, a pioneering technique [27] employing SR model to improve the disease detection by enhancing the features in LR images for subsequent processing. Similar approaches employing SR models have been adopted in contemporary research [27], [28] to identify diseases such as wheat yellow rust [29], potato late blight [30], [31], and tea leaf blight [26] in order to boost the predictive capacity of the model.

The primary contributions of the suggested research are:
A novel ViT-based backbone architecture termed DataLiteViT which employs Displaced Patch Embedding to enhance receptive fields and learn low range dependency in the small sized dataset and capture more spatial information to accurately detect disease.

The demonstration that the proposed integrated framework outperforms state-of-the-art methods in identifying diseases from deteriorated images.
The rest of the article is structured as follows. Section 2 presents literature survey. Section 3, explicates the proposed methodology. Details of the image dataset analyzed and outline of experimental implementation are presented in Sections 4 and 5. This is followed by evaluation & analyses of the experimental performance with state-of-the-art models in Section 6.Section 7 outlines ablation study and finally, section 8 offers conclusions and suggestions for future work.

II. RELATED WORK

In recent years, an enormous amount of research has been done on the automatic classification and high-accuracy detection of plant diseases. Due to the rapid evolution of deep learning, machine learning, and image processing methods, considerable advancements have been made demonstrating promising results for disease identification

Literature Review of SR
Single Image Super-Resolution (SISR) is a prominent topic within computer vision research with wide-ranging applications. A pioneering deep learning approach [32] for super-resolution (SR) reconstruction was developed by employing a three-layer convolutional neural network (CNN) surpassing traditional methods. Subsequent research achieved remarkable results in its representational capability, with emphasis on innovative model designs, such as dense blocks [33] and residual blocks [34], [35], [36].

In order to manage the increasing amount of parameters and processing power in such larger networks, recursive learning has been utilized to reduce the number of parameters by overlapping weights and using skip connections, as in DRCN [37]. Recursive learning methods require a lot of processing power as they perpetually employ the same set of parameters to get superior performance. Several attention mechanisms have also been integrated into the model as an alternative way to effectively improve restoration performance, as shown by RCAN [38] with channel attention and PAN [39] with pixel attention. However these networks have inherently complex architecture and a large number of trainable parameters.

Simultaneously, significant research has been done on developing lightweight CNN models to optimize the balance between complexity and performance in super-resolution (SR). Although SRCNN[32] was a pioneer in the CNN-based super-resolution method, it was computationally expensive. For instance, FSRCNN [40] improved upon the initial SRCNN model by substituting a highly efficient transposed convolution layer with smaller kernels for pre-upsampling. CARN-M and CARN [41] utilize both local and global cascading architectures to improve performance through effective parameter usage. IMDB [42] employed contrast-aware channel attention at the end of the block to highlight key channel features produced by successive filtering for efficient utilization of resources. The subsequent study, RFDN [43], enhances network performance by incorporating selective fusion [35] module and a feature extraction mechanism. These studies collectively underscore the substantial scope for improving the balance between model complexity and performance

Literature Review of grape leaf disease detection

Numerous studies on grape leaf disease detection have examined recent developments in deep learning techniques in an attempt to improve disease detection accuracy and efficiency. A deep learning algorithm [44] based on YOLOv3 which employs spatial pyramid pooling, with super-resolution image enhancement as a preprocessing step, has detected black rot in grape leaves with 95.79% accuracy. Another approach[45] outperformed conventional techniques by creating a precise black rot spot detection method utilizing DeepLab V3+ and a ResNet 101 backbone. Dataset enhancement using deep learning approaches and CNN-based VGG architecture [46] enabled precise grape leaf disease identification, attaining an astounding accuracy of 98.40% with hyper-parameter modification.

In another method [47], utilising CNN-SVM classifier identified five grapevine leaf species attaining 97.60% classification accuracy. This was achieved by combining feature extraction from MobileNetv2's logits layer with SVM classification utilizing various kernels and Chi-squares feature selection. Additionally, the diagnosis of Grape black measles disease was investigated [48] by using fuzzy logic with a DeepLabV3+ segmentation model based on ResNet-50.

Based on extracted attributes and the number of infections linked to the measles illness, the method divides disease severity into healthy, mild, medium, and severe categories. These studies collectively underscore a clear trend toward the adoption of advanced deep learning models to enhance the accuracy and efficacy of grape disease detection methods.

III. METHODOLOGY

The novel method employs the TSRNet[49] based Super Resolution (SR) technology to enhance a low-resolution, degraded grape leaf image into a super-resolution image. Subsequently, a DataLiteViT based classification model is utilized to identify diseases inside these improved images. This integrated TSRNet & DataLiteViT based model offers significantly superior classification performance for disease diagnosis from LR images.

Super Resolution based on TSRNet

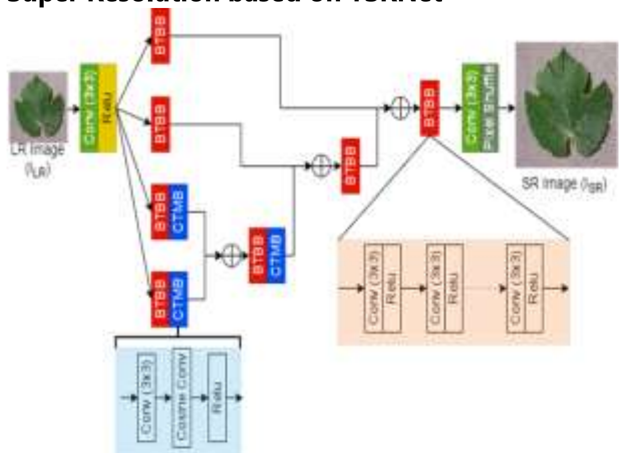

Fig. 1. Architecture of TSRNet Super Resolution method

Figure 1 shows the proposed TSRNet super-resolution (SR) architecture by integrating a tree-based backbone with cross-domain feature extraction to refine structural details.

- Tree-Guided Backbone: The network features four distinct branches composed of Binary Tree Branch Blocks (BTBB). These branches are fused sequentially to amplify the relationship between key nodes and recover complex textures.
- Cosine Transform Mechanism Block (CTMB): To prevent information loss, CTMBs use Sharpened Cosine Similarity (SCS) instead of standard dot products. This allows the model to focus on directional feature orientation rather than absolute pixel scale, improving local saliency.
- Reconstruction: A final Pixel-Shuffle layer transforms low-frequency features into high-resolution outputs.

By combining recursive tree fusion with frequency-domain insights, TSRNet achieves a superior balance between restoration accuracy and computational efficiency, consistently delivering high-fidelity results.

Classification of disease based on DataLiteViT

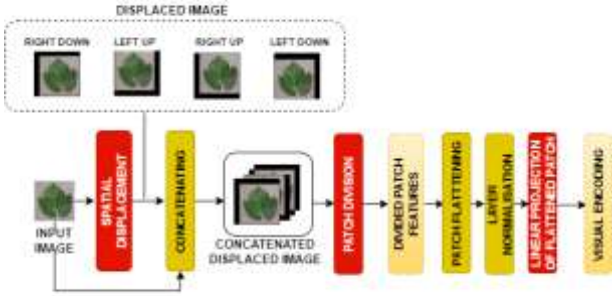

Fig. 2. Network architecture of Displaced Patch embedding mechanism

The proposed method is based on a key mechanism of Displaced Patch Embedding for improved receptive fields as depicted in figure 2. DPE leverages data augmentation by spatially displacing input image by half patch size in diagonal directions. These shifted images are then cropped to original image dimension and concatenated with the original input image. This enhances the capability of model to learn low range dependency within the data and capture more spatial information in the output embeddings. This enriched input representation is then divided into non-overlapping patches, flattened, and converted into a 1D vector series (S(i)) representation, as shown in equation 1:

$$S(i) = [i_p^1; i_p^2; i_p^3; \dots \dots ; i_p^N]$$

Where, i represents the input image with dimension h×w×c, i_p^j is the 1D representation of jth flattened concatenated patch features of displaced image i , & N indicates total number of patches (N=hw/p2), where p is patch size

These vectors series are subsequently processed by the neural network using the layer normalisation and linear projection as shown in the equation 2.

$$E(i) = LN([v_p^1; v_p^2; v_p^3; \dots \dots ; v_p^N]) * LP$$

Where, E(i) is the visual embedding output of image i, LN is the Linear normalization function and LP is the learnable linear projection used for transformation.

These encoded visual embeddings (E(i)) are then concatenated with class token and added to learnable positional embedding to incorporate positional information in the Visual Encoding layer. In the absence of a class token, the input encoded vectors are augmented solely with learnable positional embeddings.

$$\begin{aligned} &if\ v_{cls}\ is\ not\ null,\ E_{pe} = [v_{cls}; E(i)] + PE \\ &elseif\ v_{cls}\ is\ null,\ E_{pe} = [E(i)] + PE \end{aligned}$$

Where, v_cls is the class token, PE is the positional embedding & 〚 E〛_pe is the final output from the visual encoding layer.

These visual embeddings (〚 E〛_pe) are subsequently processed with a Vision Transformer (ViT) [50] like-architecture comprising of few stages of various multihead attention, layer normalization, dense, dropout, add modules in order to achieve high disease detection accuracy.

IV. DATASET

In proposed methodology widely recognised open-access PlantVillage dataset [50] comprising of greater than 54,000 sample images of plant leaves was employed. This freely accessible resource has gained popularity in recent years for exploring distinct diseases affecting diverse leaf crops. This dataset encompasses images depicting both healthy and diseased leaves, categorized into 38 categories spanning across 14 different plant crop species. Furthermore, a consistent background and homogeneous illumination were ensured when taking these images in controlled lab environments. However, the dataset is also imbalanced with varying quantities of images for various disease classes. It includes 4,062 images of grape leaves exhibiting typical symptoms.

Among these images, 1,180 were identified as having Black Rot, 1,383 as having Esca measles, 1,076 as having Leaf spot, and 423 as representing healthy leaves, all with a consistent resolution of 256 × 256 pixels. Additionally, the ground truth and its equivalent Low-Resolution (LR) grape leaf image can be observed at different scale factors (×2, ×4, and ×6) in Figure 3.

Table 1. PlantVillage grape leaf description

Type of leaf	No of scans
Black Rot	1180
Esca measles	1383
Leaf spot	1076
Healthy leaves	423
Total grape leaf images	4062

Fig. 3. Grape leaf images for scale factor ×2, ×4, and ×6

V. EXPERIMENTATION

High resolution images(IHR), or the ground truth, are downscaled by a variety of scale factors to produce LR images (ILR), which are subsequently utilized as input by the super resolution model named TSRNet. Next, using a classification model termed DataLiteViT, on these enhanced SR pictures, the disease is identified as depicted in Figure 4.

Fig. 4. Suggested approach implementing TSRNet with DataLiteViT for precise disease detection

The proposed methodology involves allocating 10% of dataset for testing, 20% for validation, and the remaining 70% for training the model. To assess the quality of the generated SR images, the loss function employed is MSE (Mean Square Error), and effectiveness of super resolution technique is measured using metrics such as SSIM (Structural Similarity Index) and PSNR (Peak Signal to Noise Ratio). Figure 5 visualizes the performance of TSRNet model and the integrated TSRNet–DataLiteViT framework by plotting training and validation curves across epochs.

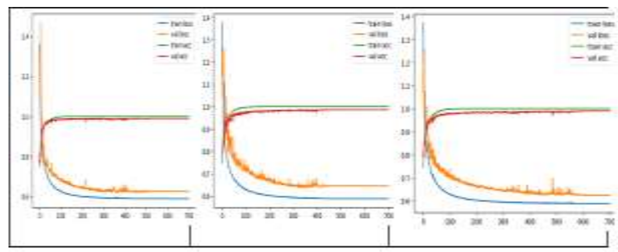

Fig. 5 Loss and accuracy curve for x2, x4, and x6 for DataLiteViT

VI. RESULTS

In this assessment, a thorough analysis is employed and the results are presented using both qualitative and quantitative methods. To effectively assess the quantitative performance of the model, a comparative analysis is conducted using two key metrics: PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) as shown in Table 2.

Table 2. Comparison table (PSNR/SSIM) for various models

Model	No. of parameters (in Millions)	×2 input size (PSNR/ SSIM)	×4 input size (PSNR/ SSIM)	×6 input size (PSNR/ SSIM)
TSRNet (2022)	2.25	30.12/0.9134	28.38/0.8416	27.89/0.756
EWSCNN(2023)[51]	0.15	30.1785/0.9042	28.446/0.8474	27. 158/0.7315
SRResCGAN (2020) [52]	1.16	29.2416/0.8913	28.695/0.8346	27.0761/0.7541
DBPN(2018) [53]	4.97	29.1849/0.8826	28.328/0.8309	26.5835/0.7227
CARN(2018) [41]	1.6	28.802/0.8927	27.006/0.8230	25.2061/0.7369
SRCNN(2014) [32]	0.057	26.8513/0.8231	25.151/0.7607	23.7014/0.7021

Table 3. Comparison of classification accuracy for various scale factors

Model	No. of parameters (in Million)	Classification Accuracy in % for scale factor ×2	Classification Accuracy in % for scale factor ×4	Classification Accuracy in % for scale factor ×6
DataLiteViT	7.1	97.90	97.12	96.69
DLVTNet(2025)[54]	1.05	96.83	94.12	91.24
MobileNetv4(2024) [55]	9.2	96.83	94.85	92.28
Eanet(2022) [56]	6.1	96.98	94.36	93.01
Gmlp(2021) [57]	19	97.29	96.57	93.63
ViT (2020) [58]	86	94.71	94.74	93.93
VGG16(2015) [59]	14	95.54	90.99	89.28

Table 3 provides a comprehensive assessment of the classification accuracy, offering a detailed comparison between the proposed DataLiteViT classification model and other state-of-the-art (SOTA) models. As observed in Table 4, the recommended approach employing TSRNet with only 2.25M additional parameters, consistently improved DataLiteViT's accuracy by more than 1% across a range of scale factors.

Table 4: Performance improvement of the proposed DataLiteViT model by using the SR images (ISR) from TSRNet as input for various scale factors

Model	No. of parameters (in Million)	Classification Accuracy in % for scale factor ×2	Classification Accuracy in % for scale factor ×4	Classification Accuracy in % for scale factor ×6
Proposed methodology	9.35	**99**	**98.71**	**98.28**
DataLiteViT	7.1	97.90	97.12	96.69

Table 5 reveals a comprehensive analysis where multiple performance evaluation metrics are employed to assess the effectiveness of various classification models specifically for a scale factor of 2. These metrics encompass a range of statistical measures, including specificity, F1 score, recall, accuracy, Matthews Correlation Coefficient (MCC) and precision [60].

Table 5: Performance comparison of the various classification models

Model	Accuracy (%)	Specificity	Recall	Precision	F1score	MCC
Proposed methodology	99	0.9990	0.9854	0.9879	0.9866	0.987
DataLiteViT	97.90	0.9960	0.9511	0.9700	0.9590	0.963
DLVTNet(2025) [54]	96.83	0.9914	0.9382	0. 9476	0.9424	0.947
MobileNetv4(2024) [55]	96.83	0.9928	0.9444	0.9549	0.9484	0.952
Eanet(2022) [56]	96.98	0.9921	0.9556	0.9649	0.9597	0.959
Gmlp(2021) [57]	97.29	0.9942	0.9557	0.9645	0.9593	0.959
ViT (2020) [58]	94.71	0.9891	0.9225	0.9327	0.9252	0.930
VGG16(2015) [59]	95.54	0.9918	0.9308	0.9406	0.934	0.913

It is clearly evident that the recommended approach employing TSRNet substantially improved the DataLiteViT model's disease detection competence. As indicated in Table 4 the DataLiteViT model managed to identify disease on LR images(x2) with 97.90% accuracy. However with the introduction of TSRNet, the accuracy increased to 99%. This demonstrates an approximately one percent improvement in classification performance with an evident increase in all key metrics like recall, F1 score etc. This implies that the suggested approach not only raised overall accuracy but also preserved an excellent balance between precision and recall.

Figure 6 illustrates the corresponding confusion matrix, which offers a graphical depiction of how the model correctly and incorrectly classifies data, aiding in a more intuitive understanding of the model's classification outcomes. This combined use of above mentioned quantitative metrics and visual representation enhances the depth of evaluation, making it more comprehensive and informative.

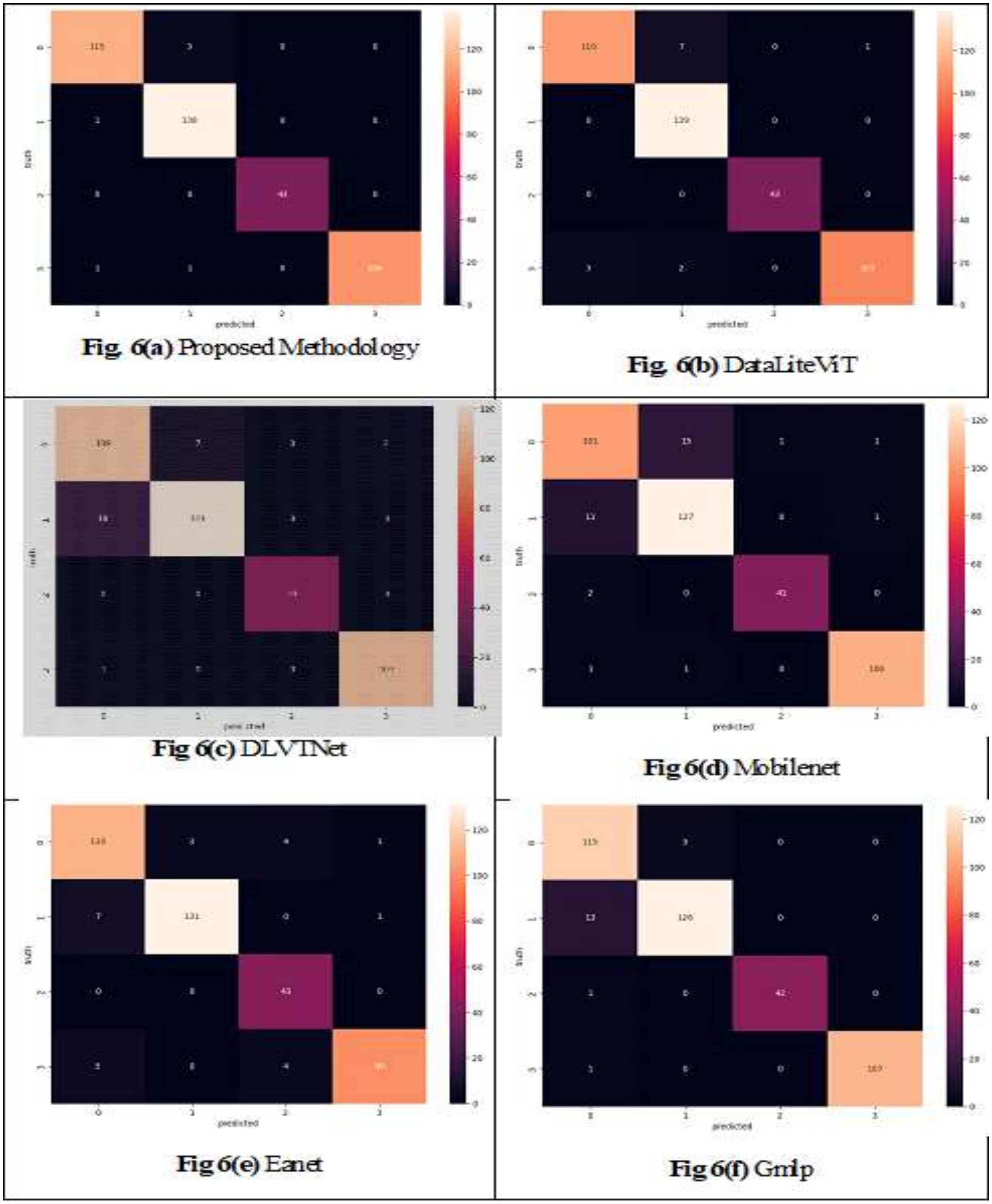

Fig.6 (a-h). Confusion matrix of the proposed methodology and various state-of-the-art models for classification of the grape plant leaf for ×2 low-resolution images.

VII. ABLATION STUDY

Table 6 Ablation experiment results for ILR (×2)

TSRNet	DPE	ViT backbone	Accuracy (%)	Specificity	Recall	Precision	F1score	MCC
✓	✓	✓	**99**	**0.9990**	**0.9854**	**0.9879**	**0.9866**	**0.987**
✓		✓	98.28	0.9894	0.9632	0.9680	0.9716	0.956
	✓	✓	97.90	0.9960	0.9511	0.9700	0.9590	0.963
		✓	97.06	0.9896	0.9732	0.9724	0.9727	0.954

Ablation study verified that the disease detection efficiency of ViT backbone was considerably enhanced by employing the proposed TSRNet and DPE modules. The quantitative results shown in Table 6 & 7 demonstrate the progressive performance increase achieved through the sequential integration of these components and also confirm their complimentary contributions to the overall architecture. The model represents the regular ViT architecture when the ViT backbone is exclusively used for classification.

Table 7 Ablation experiment results across various scale factors

TSRNet	DPE	ViT backbone	Classification Accuracy in % for scale factor ×2	Classification Accuracy in % for scale factor ×4	Classification Accuracy in % for scale factor ×6
✓	✓	✓	**99**	**98.71**	**98.28**
✓		✓	98.28	98.53	97.30
	✓	✓	97.90	97.92	96.69
		✓	97.06	97.55	96.32

Influence of DPE on model's efficiency

Table 7 describes that the ability to identify objects of the ViT model increased by 0.5% across various scale factors, when DPE was used in place of the original patch partitioning approach, demonstrating that DPE successfully improves recognition performance. Additional investigation reveals that increasing the patch size in DPE from 8 to 16 and 32, gradually increases the accuracy. Patches with smaller sizes result in more patches, splitting features that ought to be processed together and jeopardize feature integrity. On the other hand, larger patches retain more comprehensive attributes, which improves identification accuracy. The DPE method in the DataLiteViT outperformed the standard patch partitioning method when used for feature extraction.

Influence of TSRNet on model's efficiency

Table 7 demonstrates that the accuracy of DataLiteViT and the standard ViT for grape leaf images improved significantly, when TSRNet was deployed to assess classification performance on ISR across various scaling factors.

Assessment of generalisation performance

The pre-trained model's generalization ability outside of its training distribution on the PlantVillage dataset was tested using 100 grape leaf images from each of the four classes from the Grape400 dataset independently.

Grape400 dataset

Table 8 Comparison of classification accuracy for various scale factors

Model	Classification Accuracy in % for scale factor ×2	Classification Accuracy in % for scale factor ×4	Classification Accuracy in % for scale factor ×6
Proposed methodology	99.62	99.50	99.02
DataLiteViT	98.50	98.62	98.35

Table 9 Performance comparison for scale factor of ×2

Model	Accuracy (%)	Specificity	Recall	Precision	F1score	MCC
Proposed methodology	99.62	0.9983	0.9900	0.9950	0.9925	0.991
DataLiteViT	98.50	0.9908	0.9675	0.9724	0.9702	0.961

The dataset [61] of 1,600 RGB images of grape plant leaves have been separated into four classes: Healthy-leaf, Black-rot, Leaf-blight, and Black-measles. There are 400 images dedicated to each class. Table 8 and 9 compares the proposed DataLiteViT classification models on ILR and ISR across various scale factors on Grape400 dataset.

VIII. DISCUSSION

The proposed DataLiteViT model outperforms existing state-of-the-art (SOTA) models achieving accuracies of 97.90, 97.12, and 96.69 for scale factors of × 2, × 4, and × 6, respectively. The TSRNet model demonstrates impressive performance, achieving PSNR values of 30.12 dB, 28.38 dB, and 27.57dB, along with SSIM values of 0.913, 0.841, and 0.756 for scale factors of × 2, × 4, and × 6, respectively. This enhancement provided the DataLiteViT classification model with more detailed and reliable information, enabling it to produce more accurate predictions. The results demonstrate that the recommended approach employing TSRNet super-resolution with DataLiteViT classification method significantly improves overall classification performance by successfully overcoming the drawbacks of low-resolution images. DataLiteViT's accuracy progressively increased by over 1% across a variety of scale factors with just 2.25M additional parameters by utilizing SR images (ISR) as input. As a result, the classification process outperforms existing state-of-the-art (SOTA) models, yielding classification accuracies of 99.3%, 98.71% and 98.28% for scale factors of ×2, ×4, and ×6, respectively.

IX. CONCLUSION

The novel methodology presented here demonstrates its effectiveness in resolving a problem that is frequently encountered in prevalent technique by precise detection of illnesses from low-resolution images. It utilizes an innovative TSRNet-based super-resolution (SR) approach to generate enhanced resolution (ISR) images from low-resolution (ILR) images with different scaling factors, specifically ×2, ×4, and ×6. These SR images (ISR) are subsequently employed for disease detection through classification.

The study can easily be extended to encompass various other crop disease detection at different stages. Hence, it is imperative for fostering sustainable ecological development. Subsequent research endeavors can focus on the acquisition of images under varying environmental conditions and the inclusion of a broader range of leaf disease categories to enhance the accuracy of disease detection in real-world scenarios. Future research work should focus on the development of automated drones capable of swiftly assessing field conditions, thereby empowering farmers and mitigating agricultural losses.

Conflicts of Interest: The authors certify that they have NO Conflicts of Interest. Data Availability Statement: No external dataset is created and the dataset used in this research is available online for research purposes.

REFERENCES

1. P. B. Padol and A. A. Yadav, "SVM classifier based grape leaf disease detection," in 2016 Conference on Advances in Signal Processing (CASP), Jun. 2016, pp. 175–179. doi: 10.1109/CASP.2016.7746160.
2. N. Sulaiya and S. Banerjee, "A Review on Plant Leaf Disease Detection Using CNN Deep Learning Models | International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)." Accessed: Aug. 25, 2025. [Online]. Available: https://www.ijarmt.com/index.php/j/article/view/260
3. A. Subeesh and C. R. Mehta, "Automation and digitization of agriculture using artificial intelligence and internet of things," Artif. Intell. Agric., vol. 5, pp. 278–291, Jan. 2021, doi: 10.1016/j.aiia.2021.11.004.
4. S. Sharma, V. Anand, and S. Singh, "Classification of Diseased Potato Leaves Using Machine Learning," in 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), Jun. 2021, pp. 554–559. doi: 10.1109/CSNT51715.2021.9509702.
5. A. Mathew, A. Antony, Y. Mahadeshwar, T. Khan, and A. Kulkarni, "Plant disease detection using GLCM feature extractor and voting classification approach," Mater. Today Proc., vol. 58, pp. 407–415, Jan. 2022, doi: 10.1016/j.matpr.2022.02.350.
6. Bhagat M & Kumar D, "Efficient feature selection using BoWs and SURF method for leaf disease identification | Multimedia Tools and Applications." Accessed: Aug. 25, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s11042-023-14625-5
7. I. Ahmed and P. K. Yadav, "A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases," Sustain. Oper. Comput., vol. 4, pp. 96–104, Jan. 2023, doi: 10.1016/j.susoc.2023.03.001.
8. Kumar A & Patel V K, "Classification and identification of disease in potato leaf using hierarchical based deep learning convolutional neural network | Multimedia Tools and Applications." Accessed: Aug. 25, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s11042-023-14663-z
9. A. Srinivas, T.-Y. Lin, N. Parmar, J. Shlens, P. Abbeel, and A. Vaswani, "Bottleneck Transformers for Visual Recognition," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16519–16529. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2021/html/Srinivas_Bottleneck_Transformers_for_Visual_Recognition_CVPR_2021_paper.html
10. B. Neyshabur, "Towards Learning Convolutions from Scratch," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2020, pp. 8078–8088. Accessed: Aug. 25, 2025. [Online]. Available: https://proceedings.neurips.cc/paper/2020/hash/5c528e25e1fdeaf9d8160dc24dbf4d60-Abstract.html
11. C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era," presented at the Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 843–852. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Sun_Revisiting_Unreasonable_Effectiveness_ICCV_2017_paper.html
12. B. H. Ngo, N.-T. Do-Tran, T.-N. Nguyen, H.-G. Jeon, and T. J. Choi, "Learning CNN on ViT: A Hybrid Model to Explicitly Class-specific Boundaries for Domain Adaptation," presented at the Proceedings of the IEEE/CVF Conference

on Computer Vision and Pattern Recognition, 2024, pp. 28545–28554. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2024/html/Ngo_Learning_CNN_on_ViT_A_Hybrid_Model_to_Explicitly_Class-specific_CVPR_2024_paper.html
13. J. N. Mogan, C. P. Lee, M. L. Kian, M. Ali, and A. Ali, "Gait-CNN-ViT: Multi-Model Gait Recognition with Convolutional Neural Networks and Vision Transformer." Accessed: Aug. 25, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/23/8/3809
14. Y. Xie and Y. Liao, "Efficient-ViT: A Light-Weight Classification Model Based on CNN and ViT | Proceedings of the 2023 6th International Conference on Image and Graphics Processing." Accessed: Aug. 25, 2025. [Online]. Available: https://dl.acm.org/doi/abs/10.1145/3582649.3582676
15. X. Liu, H. Peng, N. Zheng, Y. Yang, H. Hu, and Y. Yuan, "EfficientViT: Memory Efficient Vision Transformer With Cascaded Group Attention," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 14420–14430. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2023/html/Liu_EfficientViT_Memory_Efficient_Vision_Transformer_With_Cascaded_Group_Attention_CVPR_2023_paper.html
16. M.-H. Guo, C.-Z. Lu, Z.-N. Liu, M.-M. Cheng, and S.-M. Hu, "Visual attention network," Comput. Vis. Media, vol. 9, no. 4, pp. 733–752, Dec. 2023, doi: 10.1007/s41095-023-0364-2.
17. A. C. Li, Y. Tian, B. Chen, D. Pathak, and X. Chen, "On the Surprising Effectiveness of Attention Transfer for Vision Transformers," Adv. Neural Inf. Process. Syst., vol. 37, pp. 113963–113990, Dec. 2024, Accessed: Aug. 25, 2025. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2024/hash/ceb45c48e29e7f49b0b47edb98e43691-Abstract-Conference.html
18. W. Wang, J. Wang, D. Quan, M. Yang, J. Sun, and B. Lu, "PolSAR Image Classification Via a Multigranularity Hybrid CNN-ViT Model With External Tokens and Cross-Attention," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 17, pp. 8003–8019, 2024, doi: 10.1109/JSTARS.2024.3384420.
19. Y. Huang, D. Wang, B. Wu, and D. An, "NST-YOLO11: ViT Merged Model with Neuron Attention for Arbitrary-Oriented Ship Detection in SAR Images - ProQuest." Accessed: Aug. 25, 2025. [Online]. Available: https://www.proquest.com/openview/89eae038c665ce3e97616732f99fbda9/1?pq-origsite=gscholar&cbl=2032338
20. J. Deng, W. Dong, R. Socher, J. Li, K. Li, and F. Li, "ImageNet: A large-scale hierarchical image database | IEEE Conference Publication | IEEE Xplore." Accessed: Aug. 28, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/5206848
21. B. Bosquet, "STDnet: A ConvNet for Small Target Detection," 2018.
22. J. Wen, Y. Shi, X. Zhou, and Y. Xue, "Crop Disease Classification on Inadequate Low-Resolution Target Images." Accessed: Aug. 25, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/20/16/4601
23. P. V. Yeswanth and S. Deivalakshmi, "ASFESRN: bridging the gap in real-time corn leaf disease detection with image super-resolution," Multimed. Syst., vol. 30, no. 4, p. 175, Jun. 2024, doi: 10.1007/s00530-024-01377-x.
24. Jorge. Torres-Sanchez, F. López-Granados, A. Isabel De Castro, and J. M. Pena-Barragan, "Configuration and Specifications of an Unmanned Aerial Vehicle (UAV) for Early Site Specific Weed Management | PLOS One." Accessed: Aug. 25, 2025. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058210
25. S. Makam, B. K. Komatineni, S. S. Meena, and U. Meena, "Unmanned aerial vehicles (UAVs): an adoptable technology for precise and smart farming," Discov. Internet Things, vol. 4, no. 1, p. 12, Sep. 2024, doi: 10.1007/s43926-024-00066-5.
26. Y. Jiang, Z. Wei, and G. Hu, "Detection of tea leaf blight in UAV remote sensing images by integrating super-resolution and detection networks," Environ. Monit. Assess., vol. 196, no.

11, p. 1044, Oct. 2024, doi: 10.1007/s10661-024-13221-w.

27. K. Yamamoto, T. Togami, and N. Yamaguchi, "Super-Resolution of Plant Disease Images for the Acceleration of Image-based Phenotyping and Vigor Diagnosis in Agriculture," Sensors, vol. 17, no. 11, p. 2557, Nov. 2017, doi: 10.3390/s17112557.
28. T. Jiang, Q. Yu, Y. Zhong, and M. Shao, "PlantSR: Super-Resolution Improves Object Detection in Plant Images." Accessed: Aug. 25, 2025. [Online]. Available: https://www.mdpi.com/2313-433X/10/6/137
29. M. H. Maqsood, R. Mumtaz, I. Ul Haq, U. Shafi, S. Mohammad Hassan Zaidi, and M. Hafeez, "Super Resolution Generative Adversarial Network (SRGANs) for Wheat Stripe Rust Classification." Accessed: Aug. 25, 2025. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7903
30. M. Sarah, M. Abdlemadjid, B. Sarah, Y. Hadjadj, and chikr el-mezouar Miloud, "Evaluating the effect of super-resolution for automatic plant disease detection: application to potato late blight detection," Multimed. Tools Appl., vol. 83, pp. 1–19, Feb. 2024, doi: 10.1007/s11042-024-18574-5.
31. P. V. Yeswanth, R. Khandelwal, and S. Deivalakshmi, "Super Resolution-Based Leaf Disease Detection in Potato Plant Using Broad Deep Residual Network (BDRN)," SN Comput. Sci., vol. 4, no. 2, p. 112, Dec. 2022, doi: 10.1007/s42979-022-01514-1.
32. C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Cham: Springer International Publishing, 2014, pp. 184–199. doi: 10.1007/978-3-319-10593-2_13.
33. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual Dense Network for Image Super-Resolution," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Residual_Dense_Network_CVPR_2018_paper.html
34. X. N. Fernando, K. Illanko, and M. Hassan, "Single Image Super Resolution using Deep Residual Learning," Nov. 03, 2023. doi: 10.20944/preprints202311.0231.v1.
35. J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual Feature Aggregation Network for Image Super-Resolution," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2359–2368. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/html/Liu_Residual_Feature_Aggregation_Network_for_Image_Super-Resolution_CVPR_2020_paper.html
36. S. Mardieva, S. Ahmad, S. Umirzakova, M. Aashik Rasool, and T. K. Whangbo, "Lightweight image super-resolution for IoT devices using deep residual feature distillation network - ScienceDirect." Accessed: Aug. 25, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950705123010912
37. J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1637–1645. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2016/html/Kim_Deeply-Recursive_Convolutional_Network_CVPR_2016_paper.html
38. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," presented at the Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 286–301. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_ECCV_2018/html/Yulun_Zhang_Image_Super-Resolution_Using_ECCV_2018_paper.html
39. H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient Image Super-Resolution Using Pixel Attention," in Computer Vision – ECCV 2020 Workshops, A. Bartoli and A. Fusiello, Eds., Cham:

Springer International Publishing, 2020, pp. 56–72. doi: 10.1007/978-3-030-67070-2_3.

40. C. Dong, C. C. Loy, and X. Tang, "Accelerating the Super-Resolution Convolutional Neural Network," in Computer Vision – ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 391–407. doi: 10.1007/978-3-319-46475-6_25.
41. N. Ahn, B. Kang, and K.-A. Sohn, "Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network," in Computer Vision – ECCV 2018, vol. 11214, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11214. , Cham: Springer International Publishing, 2018, pp. 256–272. doi: 10.1007/978-3-030-01249-6_16.
42. Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight Image Super-Resolution with Information Multi-distillation Network," in Proceedings of the 27th ACM International Conference on Multimedia, Oct. 2019, pp. 2024–2032. doi: 10.1145/3343031.3351084.
43. [43] J. Liu, J. Tang, and G. Wu, "Residual Feature Distillation Network for Lightweight Image Super-Resolution," Sep. 24, 2020, arXiv: arXiv:2009.11551. Accessed: Jul. 21, 2024. [Online]. Available: http://arxiv.org/abs/2009.11551
44. J. Zhu, M. Cheng, Q. Wang, H. Yuan, and Z. Cai, "Grape Leaf Black Rot Detection Based on Super-Resolution Image Enhancement and Deep Learning," Front. Plant Sci., vol. 12, Jun. 2021, doi: 10.3389/fpls.2021.695749.
45. H. Yuan, J. Zhu, Q. Wang, M. Cheng, and Z. Cai, "Frontiers | An Improved DeepLab v3+ Deep Learning Network Applied to the Segmentation of Grape Leaf Black Rot Spots." Accessed: Aug. 25, 2025. [Online]. Available: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.795410/full
46. A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG," Artif. Intell. Agric., vol. 6, pp. 23–33, Jan. 2022, doi: 10.1016/j.aiia.2021.12.002.
47. M. Koklu, M. F. Unlersen, I. A. Ozkan, M. F. Aslan, and K. Sabanci, "A CNN-SVM study based on selected deep features for grapevine leaves classification," Measurement, vol. 188, p. 110425, Jan. 2022, doi: 10.1016/j.measurement.2021.110425.
48. M. Ji and Z. Wu, "Automatic detection and severity analysis of grape black measles disease based on deep learning and fuzzy logic," Comput. Electron. Agric., vol. 193, p. 106718, Feb. 2022, doi: 10.1016/j.compag.2022.106718.
49. Z. Luo et al., "Fast Nearest Convolution for Real-Time Efficient Image Super-Resolution," Aug. 24, 2022, arXiv: arXiv:2208.11609. Accessed: Jul. 21, 2024. [Online]. Available: http://arxiv.org/abs/2208.11609
50. S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using Deep Learning for Image-Based Plant Disease Detection," Front. Plant Sci., vol. 7, Sep. 2016, doi: 10.3389/fpls.2016.01419.
51. Yeswanth, P. V., and S. Deivalakshmi., "Extended wavelet sparse convolutional neural network (EWSCNN) for super resolution | Sādhanā." Accessed: Aug. 25, 2025. [Online]. Available: https://link.springer.com/article/10.1007/s12046-023-02108-0
52. R. M. Umer, G. L. Foresti, and C. Micheloni, "Deep Generative Adversarial Residual Convolutional Networks for Real-World Super-Resolution," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 438–439. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2020/html/w31/Umer_Deep_Generative_Adversarial_Residual_Convolutional_Networks_for_Real-World_Super-Resolution_CVPRW_2020_paper.html
53. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep Back-Projection Networks for Super-Resolution," presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1664–1673. Accessed: Aug. 25, 2025. [Online]. Available: https://openaccess.thecvf.com/content_cvpr_2018/html/Haris_Deep_Back-Projection_Networks_CVPR_2018_paper.html

54. N. Zhang, E. Zhang, G. Qi, F. Li, and C. Lv, "Lightweight grape leaf disease recognition method based on transformer framework," Sci. Rep., vol. 15, no. 1, p. 28974, Aug. 2025, doi: 10.1038/s41598-025-13689-7.
55. D. Qin et al., "MobileNetV4 -- Universal Models for the Mobile Ecosystem," Sep. 29, 2024, arXiv: arXiv:2404.10518. doi: 10.48550/arXiv.2404.10518.
56. M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond Self-Attention: External Attention Using Two Linear Layers for Visual Tasks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 5, pp. 5436–5447, May 2023, doi: 10.1109/TPAMI.2022.3211006.
57. H. Liu, Z. Dai, D. So, and Q. V. Le, "Pay Attention to MLPs," in Advances in Neural Information Processing Systems, Curran Associates, Inc., 2021, pp. 9204–9215. Accessed: Aug. 25, 2025. [Online]. Available: https://proceedings.neurips.cc/paper/2021/hash/4cc05b35c2f937c5bd9e7d41d3686fff-Abstract.html
58. A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Jun. 03, 2021, arXiv: arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929.
59. K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Apr. 10, 2015, arXiv: arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556.
60. C. Hong et al., "Analysis of Machine Learning Algorithms for Real-Time Gallbladder Stone Identification from Ultrasound Images in Clinical Decision Support Systems," Int. J. Comput. Intell. Syst., vol. 18, no. 1, p. 73, Apr. 2025, doi: 10.1007/s44196-025-00740-8.
61. murtaza khan, "Healthy and Disease affected Leaves of Grape Plant." figshare, Oct. 13, 2020. doi: 10.6084/m9.figshare.13083890.v1.