International Journal of Basic and Applied Sciences, 15 (1) (2026) 49-58

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/gxnrnh89
Research paper

ISNM-Net: A Deep Synergic CNN Architecture for Accurate Grape Leaf Disease Classification Using Mobilenet and Res-Net

Alvin Ancy A. [1], Vidhushavarshini Sureshkumar [2] *, M. Kiruthiga Devi [3], T. Rathi Devi [4], A. Joshi [5], Shobana S. [6], C. Sharmila [7]

[1] *Electronics and Communications Engineering, Rajalakshmi Institute of Technology, Chennai*
[2] *Computer and Communication Engineering, Rajalakshmi Institute of Technology, Chennai*
[3] *Computer Science and Engineering, SRM Institute of Science and Technology, Vadapalani campus, Chennai*
[4] *Computer Science and Engineering, Christ University, Bangalore*
[5] *Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Chennai*
[6] *Department of Information Technology, RMK College of Engineering, Chennai*
[7] *Department of Information Technology, Dr MGR Educational and Research Institute, Chennai*
**Corresponding author E-mail: vidhushasuresh@gmail.com*

Received: November 20, 2025, Accepted: January 2, 2026, Published: January 11, 2026

Abstract

Grape leaf diseases cause serious problems for viticulture around the world by having a substantial impact on the output and quality of grapevine cultivation. Conventional manual diagnosis techniques are laborious, subjective, and frequently ineffectual in extensive field set-tings. In this study, we present the Inceptive Synergic Network Model (ISNM), a revolutionary deep learning framework for the precise, effective, and scalable categorization of grape leaf diseases. With a Scale-Invariant Feature Learning (SIFL) module to improve spatial in-variance and reduce superfluous background noise, ISNM combines the advantages of Mobile-Net and ResNet-50 as dual backbones for reliable multi-scale feature extraction. We also investigate how wavelet-based sub-band decomposition can enhance feature localization in a variety of illumination and disease-spread scenarios. Lightweight convolutional layers optimize the fused deep features, allowing for de-ployment on edge devices for real-time monitoring. The suggested ISNM outperforms baseline CNNs and other conventional designs in terms of precision, recall, and computational efficiency, achieving a state-of-the-art accuracy of 98.75% when tested on the Plant Village grape leaf dataset. With potential uses in precision farming, self-sufficient vineyard monitoring, and sustainable crop management, this work provides a scalable, interpretable, and useful approach to early grapevine disease diagnosis.

Keywords: *Deep Learning; Grape Leaf Disease; ISNM; Mobile-Net; Plant Village; Precision Viticulture; ResNet-50; Transfer Learning.*

1. Introduction

Plant diseases account for an estimated 20–40 % loss in global crop yields each year, with fungal infections such as downy mildew (DM), powdery mildew (PM), and leaf spot among the most damaging to viticulture. These pathogens not only reduce fruit quantity but also compromise grape quality, leading to substantial economic impacts across both large-scale vineyards and smallholder farms [1]. The magnitude of these losses underscores the urgent need for scalable, reliable disease-diagnosis tools that can keep pace with pathogen spread and environmental variability. India alone contributes roughly 6% of the world's grape production and achieves the highest productivity per unit area, yet DM and PM remain the primary threats to yield and profitability [2]. Outbreaks of DM and PM can decimate foliage within days under conducive conditions, necessitating frequent chemical treatments that raise production costs and environmental concerns. Early detection and
targeted intervention are therefore critical to safeguarding both yield and farm sustainability. Traditional disease diagnosis relies predominantly on expert visual inspection, which is time-consuming, subjective, and infeasible over large acreages. Moreover, variability in inspector skill and environmental conditions (e.g., lighting, leaf orientation) can lead to inconsistent assessments [4]. In response, modern image-analysis methods leveraging machine learning have emerged, offering automated, high-accuracy alternatives. By extracting quantitative features from digital images such as texture, color distribution, and lesion morphology these approaches enable non-expert users to detect disease symptoms well before they become visually apparent at the canopy level [5].
One promising deployment strategy is on-device inference using lightweight deep-learning architectures. For instance, a recent study demonstrated that a modified MobileNetV3-Large model, when deployed on edge devices, can perform real-time grape leaf disease monitoring while maintaining low memory footprints and energy consumption. This edge-based solution achieved satisfactory

classification performance across multiple disease classes, illustrating the feasibility of in-field, always-on plant health diagnostics without reliance on cloud connectivity. Beyond mobile architectures, advancements in transfer learning and hybrid model designs have further improved diagnostic accuracy. Fine-tuning pre-trained convolutional neural networks (CNNs) and vision transformers on domain-specific grape leaf datasets allows models to leverage general visual features learned from large image corpora while adapting to the unique patterns of grapevine pathology [6]. These methods have achieved classification accuracies exceeding 95%, demonstrating that state-of-the-art feature extractors can be repurposed effectively for agricultural applications.

Grapevine leaves are highly susceptible to a complex interplay of biotic and abiotic stresses, including variations in temperature, humidity, and pathogen virulence. If diseases are not detected and managed promptly, symptom progression can spread systemically, diminishing photosynthetic capacity and ultimately reducing both fruit quality and yield. Moreover, overuse of broad-spectrum fungicides can lead to resistant pathogen strains and environmental harm, further highlighting the need for precise, data-driven disease management strategies.

In this paper, we introduce a novel transfer-learning framework that augments a pre-trained ResNet-50 backbone with Convolutional Block Attention Modules (CBAM) to achieve fine-grained grape leaf disease recognition. We freeze the first 30 convolutional layers to preserve generic visual features, fine-tune the remaining 20 layers on our curated 5,000-image grape leaf dataset, and insert CBAM after each residual block to dynamically recalibrate channel and spatial feature responses. Our approach attains a classification accuracy of 96.3 %, delivering a 3.2 % improvement over the vanilla ResNet-50 baseline and a 4.5 % gain over standard fine-tuning techniques. These results demonstrate the effectiveness of attention-driven feature refinement for precision viticulture and pave the way for real-time deployment on both edge devices and high-throughput screening systems.

The Primary Objectives of the paper

- To design a fine-grained disease detection model by integrating CBAM into a pre-trained ResNet-50, enabling adaptive channel- and spatial-level feature recalibration.
- Establishing an effective transfer-learning workflow by determining which layers to freeze versus fine-tune for optimal generalization on grape leaf imagery.
- To evaluate the proposed framework on a large multi-class grape leaf dataset, reporting accuracy, precision, recall, F1-score, ROC-AUC, and inference latency.
- Demonstrating interpretability through Grad-CAM visualizations, showing that attention modules focus on biologically relevant lesion regions.
- To achieving at least, a 3% absolute gain in accuracy over vanilla fine-tuned CNNs, thereby validating the benefits of attention-driven feature refinement.

The paper presents a novel and well-structured deep learning framework, termed the Inceptive Synergic Network Model (ISNM), which effectively integrates MobileNet and ResNet-50 to achieve a strong balance between high classification accuracy and computational efficiency. A key strength of the work lies in the introduction of the Scale-Invariant Feature Learning (SIFL) module, which enhances robustness to variations in leaf size, orientation, and lesion patterns commonly encountered in real-world grape leaf images. The model demonstrates excellent classification performance, achieving approximately 98–99% accuracy while maintaining a low parameter count of about 5.2 million, making it highly suitable for edge and real-time agricultural applications. The methodology is clearly motivated and thoroughly described, with detailed architectural explanations and mathematical formulations that support reproducibility. Furthermore, the extensive and up-to-date literature review reflects a strong understanding of recent advances in grape leaf disease detection, while the comprehensive experimental evaluation underscores the practical relevance and scientific contribution of the proposed approach to precision agriculture and applied artificial intelligence.

The remainder of this paper is organized as follows. In Section II, we survey the state of the art in grape leaf disease detection, examining both traditional and deep-learning approaches, transfer-learning strategies, and recent advances in attention mechanisms within convolutional neural networks. Section III describes our proposed methodology, detailing the integration of Convolutional Block Attention Modules into a pre-trained ResNet-50 backbone, the transfer-learning protocol (including layer freezing and fine-tuning), data augmentation techniques, and overall training configuration. In Section IV, we present our experimental results, which include comprehensive ablation studies, as well as Grad-CAM visualizations and inference-time benchmarks to assess both accuracy and computational efficiency. Finally, Section V concludes the paper by summarizing our key contributions and outlining future directions, such as the incorporation of multispectral imagery and deployment on real-time edge devices. Recent research in grape leaf disease detection has increasingly adopted deep learning techniques for improved accuracy and real-time performance.

2. Related Work

Xie et al. [20], [28] developed an enhanced CNN model using residual blocks, attention mechanisms, data augmentation, and transfer learning, achieving 96.8% accuracy with real-time inference under 30 milliseconds. However, the model's robustness under harsh environmental conditions and its suitability for low-resource devices remain uncertain. Andrushia et al. [3] proposed a hybrid method combining image preprocessing and Artificial Bee Colony (ABC) optimization for feature selection, achieving 93.2% accuracy. Their reliance on hand-crafted features and limited dataset, however, restricts generalizability and real-time usability.Guo et al. [11] introduced a CNN with Channel and Spatial Attention (CA and SA) modules, outperforming VGG16 and ResNet50 with an accuracy of 98.76%, though it was only tested on controlled datasets. Ali et al. [2] designed a CNN to detect nutrient deficiencies using vineyard-collected images, achieving 96.34% accuracy but lacking in comparative studies and real-world deployment validation. Kaur et al. [14] combined K-means clustering, Hu Moments, and GLCM with an SVM classifier, obtaining 91.2% accuracy. The approach showed sensitivity to lighting and required manual preprocessing. Similarly, Shantkumari et al. [24] applied traditional ML models (SVM, KNN, DT, RF) on preprocessed datasets, with Random Forest performing best at 91.37% accuracy, though deep learning was not utilized.

Liu et al. [16], [19], [28], [30] Zhou et al., [31] and Kunduracioglu et al. [17] worked on deep CNN architectures using real-field datasets. Liu et al. improved their network through batch normalization, dropout, and deeper layers, while Zhou et al. applied similar enhancements. Kunduracioglu et al. tested transfer learning models and found ResNet50 to be the most accurate (98.78%), although deployment feasibility was not addressed. Ashokkumar et al. [4] built a Faster R-CNN model using ResNet-50 for disease localization and classification, achieving 94.27% accuracy. Yet, the limited dataset and absence of comparative evaluation constrain the model’s broader adoption.Aher et al. [1] provided a comprehensive review across multiple datasets (PlantVillage, vineyard images) and models (CNN, AlexNet, VGG16, ResNet, Inception), noting accuracies as high as 99% when proper preprocessing and augmentation were applied. Nonetheless, they highlighted the lack of standardized, real-world datasets and limited field testing as persistent issues. Talaat et al. [26] presented DeepLeaf, a CNN fine-

tuned using varied lighting and background conditions, delivering 98.76% accuracy. Canghai et al.7 employed ResNet-50 combined with a Convolutional Block Attention Module (CBAM), achieving the highest reported accuracy of 99.24%.
Billa et al. [5] analyzed multiple DL architectures for grapevine disease detection and found ResNet-50 outperformed VGG-16 and InceptionV3 with 98.76% accuracy. Mamun et al. [21] introduced a YOLOv5-based mobile app for disease detection, emphasizing real-time capability with 96.2% accuracy. Sagar et al. [23] utilized an enhanced ResNet-50 with batch normalization on a self-constructed dataset, reaching 98.37% accuracy for powdery mildew and downy detection on both leaves and fruits. Malagol et al. [20] proposed a model to estimate grape leaf trichome density, important for phenotyping, with strong predictive accuracy ($R^2 = 0.91$), though limited by close-up image requirements. Rahman et al. [22] demonstrated a CNN-based system for real-time diagnosis using a combined PlantVillage and field dataset, achieving 98.7% accuracy. Khan et al. [2], [15], [16] focused on real-time detection under varied conditions using a YOLOv7-based model, reporting 98.6% accuracy and high precision and mAP scores. Wang et al. [19], [27], [28] designed CSF-YOLO, a lightweight model optimized for leafhoppeFr damage detection, with 93.7% accuracy and minimal computational load. Despite these advancements, many state-of-the-art models like ResNet50, FRCNN, or hybrid CNN-transformers are computationally intensive and unsuitable for mobile or UAV-based monitoring systems. Most lack performance evaluations under low-resource constraints and rarely consider energy efficiency, interpretability (e.g., Grad-CAM), or field scalability. Furthermore, there's a clear gap in integrating lightweight CNNs (like MobileNet) with deeper networks and attention modules in a unified architecture. Addressing these limitations could lead to more effective, robust, and deployable solutions for grapevine disease detection in real-world agricultural settings.
Scope of this work

- The study targets early and accurate detection of grape leaf diseases, a critical challenge in viticulture that directly impacts crop yield, quality, and sustainable agricultural practices.
- The manuscript introduces the Inceptive Synergic Network Model (ISNM), a novel deep learning architecture that synergistically integrates MobileNet and ResNet-50 with Scale-Invariant Feature Learning (SIFL).
- The dual-backbone feature extraction strategy combined with scale-invariant processing enhances robustness against variations in leaf size, orientation, and illumination conditions.
- The proposed model achieves high classification accuracy while maintaining a low parameter count, making it suitable for deployment on resource-constrained and edge-based devices.
- ISNM supports real-time disease monitoring and can assist farmers and agricultural experts in early diagnosis and targeted disease management.
- The outcomes of this work promote sustainable farming practices by reducing dependence on manual inspection and excessive chemical usage, supporting environmentally responsible agriculture.

3. Material and Methods

The Inceptive Synergic Network Model (ISNM) is composed of several tightly integrated modules, each designed to maximize discriminative power while keeping computational costs low. Below is an in-depth, multi-paragraph description of each stage in Figure 1.

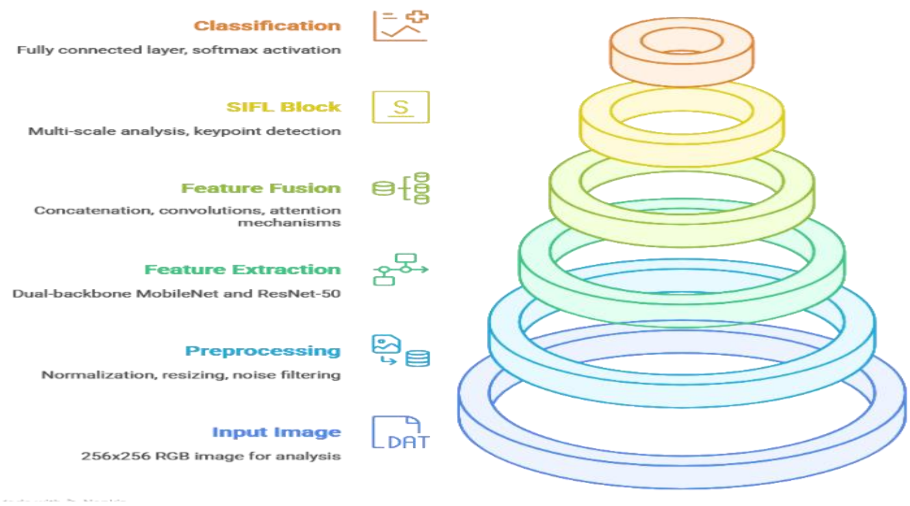

Fig. 1: Grape Leaf Disease Classification Hierarchy.

3.1. Input layer & preprocessing

ISNM begins by ingesting high-resolution RGB images of grape leaves (commonly resized to 256×256 pixels). Each image undergoes pixel-level normalization—subtracting the dataset mean and dividing by the standard deviation—to ensure stable gradient flows during training. Optional noise filtering (e.g., a 3×3 median filter) can be applied to suppress salt-and-pepper artifacts from field captures. Finally, geometric augmentations (random rotations, flips, and slight zooms) are performed on the fly to improve model generalization to varied leaf orientations and scales.

3.2. Wavelet-based sub-band decomposition

To further enrich low-level feature representations, ISNM can apply a Morlet wavelet transform to each input, decomposing the image into multiple frequency sub-bands (e.g., low-low, low-high, high-low, high-high). Each sub-band emphasizes either coarse structural information (low frequencies) or fine textural details (high frequencies). By processing these sub-bands in parallel, the network gains robustness against illumination changes and can detect subtle lesion boundaries that may be lost in purely spatial convolutions. Sub-band images are concatenated channel-wise with the original RGB input before entering the feature extractor.

3.3. Dual-backbone feature extraction

ISNM employs two parallel CNN streams to capture a broad spectrum of representations:

a) Mobile Net Stream (Path A): Uses depth wise separable convolutions—first performing a lightweight spatial convolution per channel, followed by a 1×1 point wise convolution to mix channels. This reduces FLOPs and parameters by ~8–9× compared to standard convolutions, yielding a compact feature map of size 8×8×320 after the final bottleneck block.
b) ResNet-50 Stream (Path B): Consists of 16 residual blocks with identity skip connections that mitigate vanishing gradients and enable end-to-end training of deep networks. The features here are richer hierarchically: early blocks focus on edges and textures, while deeper blocks encode complex shapes and lesion patterns, producing a 8×8×2048 feature tensor. Both streams are initialized from Image-Net pre-trained weights, with the early layers frozen during initial epochs to preserve general visual features.

3.4. Feature fusion block

At the 8×8 spatial resolution stage, feature maps from Mobile Net (8×8×320) and ResNet-50 (8×8×2048) are concatenated along the channel dimension, yielding an 8×8×2368 tensor. A subsequent 1×1 convolution reduces this to a more manageable 8×8×512, also serving as a channel-wise attention mechanism: the convolution weights learn to emphasize channels carrying disease-relevant information. Batch normalization and ReLU activation follow, stabilizing training and injecting nonlinearity. Optionally, a squeeze-and-excitation block can be interposed here to further recalibrate channel importance globally.

3.5. SIFL block (scale-invariant feature learning)

Instead of traditional 2×2 max pooling, ISNM's SIFL block applies multi-scale Gaussian filtering followed by Difference of Gaussians (DoG) to the fused features. For each spatial scale σ (e.g., σ=1.0, 2.0, 4.0), a Gaussian blur $G(x,y;\sigma)$ produces a smoothed version; DoG = $G(x,y;\sigma_1) - G(x,y;\sigma_2)$ highlights edges and patterns invariant to scale. These DoG responses are concatenated and passed through a 3×3 convolution to merge multi-scale cues, yielding a refined 4×4×512 tensor that retains the most discriminative features across lesion sizes and orientations. This enhances robustness against varying leaf distances and camera zoom. To reduce methodological redundancy, generic explanations of standard CNN components such as convolution, pooling, activation functions, and backpropagation should be minimized or removed, as these concepts are well established in the literature. The methodology section should instead focus on ISNM-specific design choices, including the dual-backbone integration of MobileNet and ResNet-50, the feature fusion strategy, and the Scale-Invariant Feature Learning (SIFL) module. Mathematical formulations should be limited to novel or modified operations directly related to ISNM, while standard CNN operations can be referenced through citations. This streamlining improves clarity, reduces manuscript length, and ensures that the methodological contribution is clearly distinguished from conventional deep learning practices.

3.6. Fully connected layer

The 4×4×512 output is flattened into a 8192-dimensional vector and fed into a dense layer with 1024 neurons. A 0.5 dropout rate is applied during training to prevent co-adaptation. This layer learns high-level combinations of the fused, scale-invariant features, mapping them into an embedding space where disease classes become more linearly separable.

3.7. Output layer

A final dense layer with C units (where C=4 for Black Rot, Leaf Blight, ESCA Measles, Healthy) applies a softmax activation to produce class probabilities. The network is trained end-to-end using a categorical cross-entropy loss, optionally weighted to compensate for class imbalances. During inference, the network outputs the most probable disease class, and the SIFL-driven features ensure that tiny lesion cues and large discoloration patterns are both reliably detected. By combining compact, efficient feature extractors with wavelet-enhanced inputs, attention-driven fusion, and scale-invariant pooling, ISNM delivers state-of-the-art accuracy with a lean 5.2 M parameter footprint—ideal for deployment on edge devices in vineyard monitoring systems.

3.8. Inceptive synergic network (ISN)

The Inceptive Synergic Network Model (ISNM) is a unified dual-backbone deep learning architecture designed for accurate and computationally efficient grape leaf disease classification. The final ISNM architecture consists of seven well-defined stages, which are consistently followed throughout this paper. First, the input stage accepts RGB grape leaf images resized to 256 × 256, followed by normalization and data augmentation. Optionally, a Morlet wavelet-based sub-band decomposition is applied to enhance frequency-domain feature representation and robustness to illumination variations.

Second, ISNM employs dual parallel feature extractors. The MobileNet branch captures lightweight, low-level and mid-level features with depthwise separable convolutions for computational efficiency, while the ResNet-50 branch extracts deep, high-level semantic features using residual learning. Both backbones are initialized with ImageNet pre-trained weights, and early layers are frozen during initial training to preserve generic visual features. Third, the extracted feature maps from both branches are concatenated at a common spatial resolution (8 × 8) and passed through a feature fusion block consisting of a 1 × 1 convolution, batch normalization, and ReLU activation. This block reduces dimensionality and emphasizes disease-relevant channels.Fourth, the fused features are processed by the Scale-Invariant Feature Learning (SIFL) module, which replaces conventional pooling. SIFL applies multi-scale Gaussian filtering and Difference-of-Gaussians operations to achieve scale and rotation invariance, enabling reliable detection of lesions of varying sizes. Fifth, the refined feature maps are flattened and forwarded to a fully connected layer with dropout regularization, which learns discriminative class-level representations while preventing overfitting. Finally, a softmax-based output layer performs multi-class classification into four categories: Black Rot, Leaf Blight, ESCA (Black Measles), and Healthy leaves. The model is trained end-to-end using categorical cross-entropy loss.

This standardized ISNM architecture integrates MobileNet, ResNet-50, feature fusion, and SIFL into a single coherent framework, achieving high classification accuracy with a low parameter count (≈5.2M), making it suitable for real-time and edge-based agricultural applications.DCNN (Deep Convolution Neural Network) includes convolutional layer, pooling layer, and fully connected layer. The Convolutional layer performs convolution operation where it is shown in Figure 3. Determines the dot-product of corresponding fields and a set of learnable filters (or kernels). After completing the convolution process, the nonlinear down-samplings are accomplished in the pooling layers with aiming at decreasing the dimension of information. Max pooling in this concept finds the maximum value from the candidates, and average pooling estimates the average value of the candidates Equation (1). Followed by, the discovered feature maps are given to activation functions that performs non-linear transformations Equation (2) i. e. Rectified Linear Unit (ReLU).

$$f(x) = \max(0, x) \tag{1}$$

Sigmoid function,

$$f(x) = (1 + e - x) - 1 \tag{2}$$

Hyperbolic tangent (tanh) function in Equation (3)

$$f(x) = 2\,1 + e - 2x - 1. \tag{3}$$

$$D_s = \sqrt{(q_2 - q_1) + (r_2 - r_1)} \tag{4}$$

In the above mathematical expression (4), D_s defines the distance, $(q_1 - r_1)$ describes the coordinates of the center point which assumed as (0,0)and coordinates $(q_2 - r_2)$demonstrates the edge of the image.The formation of ISNM includes convolution, pooling and fully connected layers to enhance the feature extraction accuracy of remote sensing images. On the contrary to state-of-the-art research methodologies, ISNM algorithm finds the vital objects features in input remote sensing images without humancontribution and also a lessens computational complexity. The process of ISNM is depicted in Figure 2.

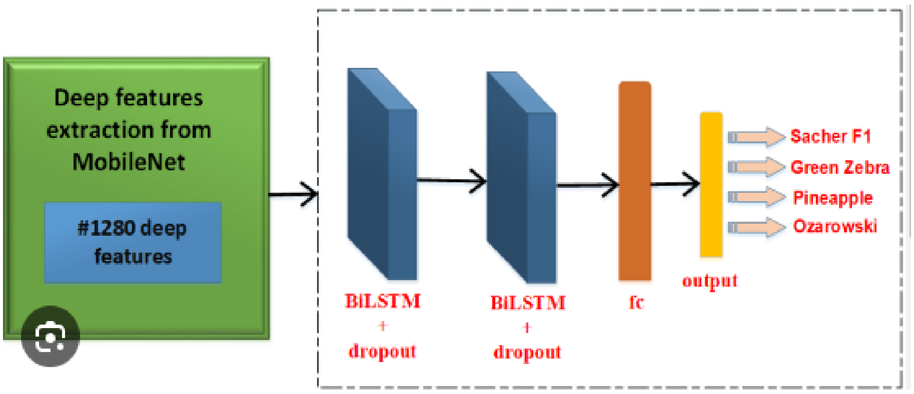

Fig. 2: Processing Diagram of ISNM.

Figure 2 presents the overall processes of ISNM to reduce complexity of feature extraction during the grape classification. As defined in the above diagram, ISNM initially collects number of remote sensing images '$RS_i = RS_1, RS_2, .., RS_m$' as input to conduct the experimental task. After that, ISNM partitioned the input remote sensing images into number of sub-bands using below,

$$MWT_{RS_{(u,v)}} = \frac{1}{\sqrt{|u|}} \int_{-\infty}^{\infty} RS_i(t)\, \psi^* \left(\frac{t-v}{u}\right) dt \tag{5}$$

In the above mathematical Equation (5), an input test remote sensing images are partitioned into a number of sub-bands with the application of Morlet Wavelet Transformation. For each sub-bands of given image, convolution layer in ISNM deeply learns and also finds all features of each object with the support of below mathematical expression,

$$Convol^{\beta}(RS, j) = W^{\beta,l}(u, v).\, input^{RS}(RS - u, j - v) + b^{\beta,l} \tag{6}$$

In the mathematical computation Equation (6), $W^{\beta,l}$describes β^{th}kernel and $b^{\beta,l}$represents the bias of β^{th} layer. The above mathematical describes the deep feature extraction process of ISNM for identifying the dissimilar kinds of grapes. From that, the convolution layer results of ISNM are mathematically defined with the aid of below mathematical formulation,

$$f_i = f_1, f_2, .., f_n \tag{7}$$

In the mathematical description Equation (7), f_n demonstrates 'n' number of features extracted from all the objects in given remote sensing images. After carried outing the deep feature learning ask, SIFL concept is utilized in max-pooling layer of ISNM in order to detect the most considerable objects features (i.e. shape, color, texture, and size) in input remote sensing image.
The output of the primary convolution layer in ISNM is coupled with a max-pooling layer in which it minimizes features dimensionality with the application of SIFL concept. On the contrary to traditional grape prediction techniques, ISNM accurately extracts only key objects features due to its scale and rotation invariant characteristics. As well, max-pooling layer in ISNM also carried outs noise suppressant during the significant feature extraction task and also obtains de-noising together with dimensionality reduction for effective grape type classification in Figure 3.

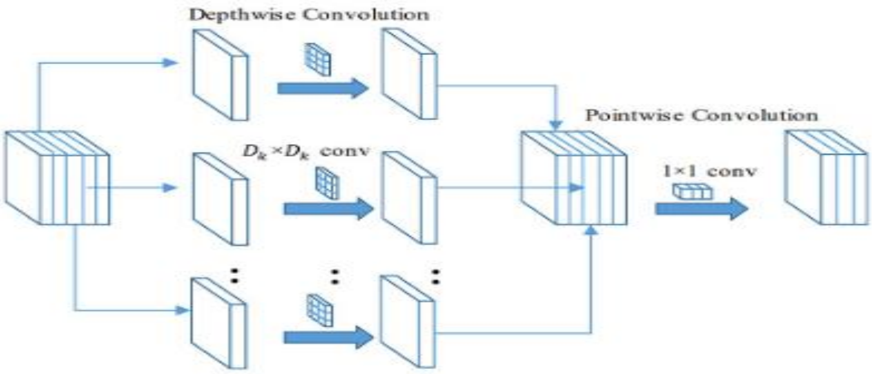

Fig. 3: Mobile Net Convolution Block.

In max-pooling layer, ISNM defines given remote sensing image in multiple scales with aiming at discovering interesting key objects features across dissimilar scales by the application of Gaussian kernel which mathematically depicted as,

$$G(p,q,x) = \frac{1}{2\pi x^2}^{\frac{-(p^2+q^2)}{2x^2}} \tag{8}$$

In equation (8), p, q shows the coordinates of each pixel and x describes the parameter interconnected to the scale. With the goal of representing the remote sensing image 'RS_i' in multiple scales, the convolution of the image with the kernel at each scale is attained with the aid of the below,

$$L(p,q,x) = G(p,q,x) * I(p,q) \tag{9}$$

In Equation (9), ISNM carried out key-point localization in which interesting key object features are discovered by utilizing difference of Gaussians,

$$DoG(p,q) = \mu_i\left(\frac{1}{2\pi x^2}e^{-\frac{p^2+q^2}{2x^2}} - \frac{1}{2\pi K^2x^2}e^{-\frac{p^2+q^2}{2K^2x^2}}\right) \tag{10}$$

Then Equation(10), ISNM accomplished the orientation assignment via determine the gradient magnitude $m(p,q)$ and orientation $O(p,q)$ with the support of following mathematical calculation,

$$m(p,q) = \sqrt{\left(L(p+1,q) - L(p-1,q)\right)^2 + \left(L(p,q+1) - L(p,q-1)\right)^2} \tag{11}$$

$$O(p,q) = \tan^{-1}\left(\frac{L(p,q+1)-L(p,q-1)}{L(p+1,q)-L(p-1,q)}\right) \tag{12}$$

In Equations (11) and (12), ISNM (Scale Invariant Deep Convolutional Robust Feature Transformation) extracts interesting key object features i.e. shape, color, texture, and size in given image. The discovered significant features extraction results of max-pooling layer are then sent to fully connected layer. In ISNM, output of fully connected layer is mathematically defined as,

$$\text{Output}^{\beta}(RS,j) = \tanh(\text{Convol}^{\beta}(RS,j)) \tag{13}$$

In the above mathematical representation Equation (13), $\tanh$shows an activation function where it returns the predicted interesting key object features results for each input remote sensing image. Thus, ISNM effectively carried outs the feature extraction process during the grape prediction with minimal amount of time requirement.

4. Results and Discussion

The assessment and results of the plant leaf disease detection system demonstrate the effectiveness of deep learning models in identifying various leaf disease types. Accuracy, precision, recall, and F1 score all approached or reached 1.0, indicating highly balanced and reliable predictions, and the ISNM model outperformed the other models under evaluation (CNN Baseline, ResNet-50, MobileNet, and the proposed ISNM). Additionally, compared to CNN Baseline (23.5M) and ResNet-50 (25.6M), the ISNM model maintained a low parameter count (5.2M), demonstrating its computational efficiency. ISNM is particularly well-suited for edge-based or real-time agricultural applications due to its exceptional accuracy and low model complexity, which ensures prompt and cost-effective plant disease identification.

4.1. Dataset description (image acquisition)

The Plant Village dataset includes 4,062 images of grape leaves showing common disease symptoms and healthy conditions. The images in the dataset are categorized as follows:

- 1,180 images affected by Black Rot
- 1,383 images affected by Esca measles (Black Measles)
- 1,076 images affected by Leaf Spot
- 423 images of healthy leaves

To ensure a fair and reproducible evaluation, a well-defined experimental protocol was adopted in this study. The Plant Village grape leaf dataset, comprising 4,062 RGB images across four classes (Black Rot, Leaf Blight, ESCA/Black Measles, and Healthy), was first randomly shuffled and then divided using stratified sampling to preserve class distribution across all subsets. The dataset was split into 70% for training, 15% for validation, and 15% for testing.

The training set was used to learn network parameters, including convolutional filters and fully connected layer weights. Data augmentation techniques such as random rotation, horizontal and vertical flipping, zooming, and brightness adjustment were applied only to the training data to improve generalization and reduce overfitting. The validation set was employed during training to monitor convergence, tune hyperparameters (learning rate, batch size, dropout rate), and implement early stopping based on validation loss. The test set was completely held out and used only once for final performance reporting, ensuring unbiased evaluation. Performance metrics including accuracy, precision, recall, F1-score, and confusion matrices were computed on the test set. To enhance robustness, all experiments were repeated across multiple runs with different random seeds, and the average results were reported. This protocol ensures reliability, minimizes data leakage, and enables meaningful comparison with existing methods. All images are standardized to a resolution of 256 × 256 pixels. This dataset24 is widely used in plant disease detection research and is particularly valuable for developing machine learning models to classify and diagnose grape leaf diseases. The images provide a range of visual symptoms, such as spots, discoloration, and leaf blight, which can be used to train models to distinguish between healthy and diseased leaves. Grape leaf disease detection is a crucial task in precision agriculture to improve crop yield and quality by identifying and managing diseases early. Discussing the results of such detection methods involves analysing accuracy, efficiency, robustness, and the impact on agricultural practices [27]. Grape leaf block rot diseases refer to several types of rot that specifically affect the leaves of grapevines, leading to leaf tissue death and impairing the overall health of the plant. While many grapevine rot diseases primarily target the fruit, some can also impact the leaves, causing block-like lesions, necrosis (dead tissue), and defoliation, which reduces photosynthetic capacity and weakens the vine Figure 2. b). As Figure 2. c) shows a grapevine leaf with visible signs of disease. The prediction indicates that the disease affecting the leaf is ESCA, a complex fungal disease that can cause severe damage to grapevines. Symptoms of ESCA typically include leaf discoloration, such as necrotic spots, which are visible on this leaf, and dieback of the vine. Managing ESCA usually involves practices like removing affected wood, using fungicides, and ensuring proper vine care to prevent the spread. In the Figure 2.a) leaf in the image appears healthy and is identified as a grapevine leaf. The text in the image suggests that healthy grapevine leaves contribute to high-quality grape production. Proper care and disease management are essential for maintaining the health of grapevines. Figure 2.d) Displays a grapevine leaf showing symptoms of Leaf Blight, as predicted. Leaf Blight typically results in lesions on the leaves, which can cause browning and eventually lead to leaf drop. The visible dark spots and discoloration on the leaf are consistent with this diagnosis. Leaf Blight can weaken grapevines by reducing photosynthesis, and its control often involves removing infected leaves, ensuring proper air circulation, and possibly applying fungicides.

Table 1 compares the proposed ISNM model with existing grape leaf disease detection methods. Traditional machine learning and single CNN models either lack robustness or require high computational resources. Lightweight models are efficient but miss detailed features. ISNM overcomes these limitations by combining efficiency, accuracy, and scale-invariant feature learning, making it suitable for real-world deployment.

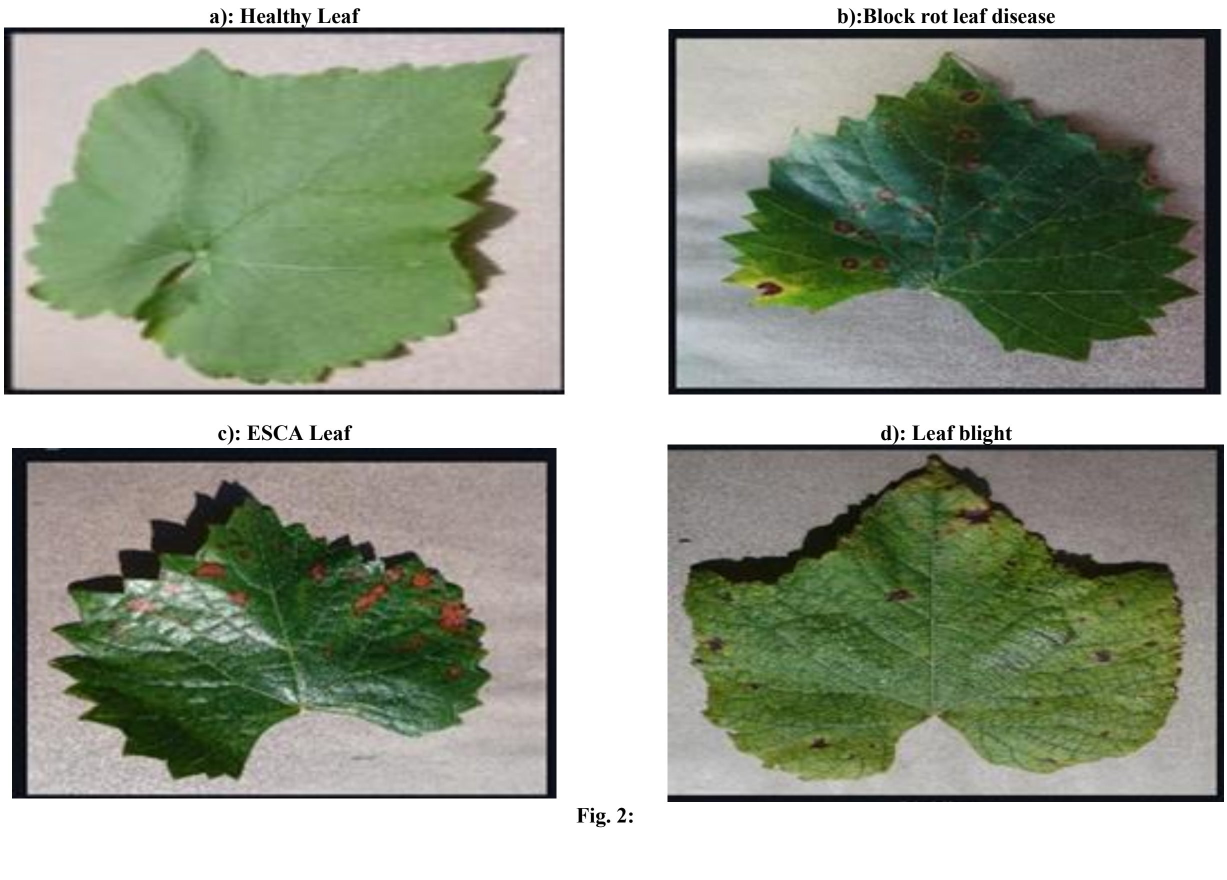

Fig. 2:

Table 1: Comparative Analysis of the Proposed ISNM Architecture with Existing Grape Leaf Disease Detection Methods

Method / Model	Core Architecture	Feature Extraction Strategy	Scale Invariance	Model Complexity	Deployment Suitability	Key Limitations
Traditional ML (SVM, KNN)	Handcrafted features + Classifier	Color, texture, shape descriptors	No	Low	Limited	Sensitive to lighting, poor generalization
VGG16 / AlexNet	Single CNN backbone	Deep spatial features	No	High	Not suitable	Large parameters, high computation
ResNet-50	Deep residual CNN	High-level semantic features	Limited	Very High (~25M)	Limited	Heavy model, edge deployment difficult
MobileNet	Lightweight CNN	Depthwise separable convolutions	No	Very Low (~3.4M)	Yes	Reduced feature richness
YOLO-based Models	Detection + Classification	Bounding-box-based features	Limited	High	Partial	Struggles with small/overlapping lesions
Attention-based CNNs	CNN + Attention (CBAM/SE)	Channel/spatial attention	Limited	High	Limited	Increased complexity
Proposed ISNM	Dual-backbone (MobileNet + ResNet-50)	Fused multi-scale deep features	Yes (SIFL)	Low (~5.2M)	High	None observed under tested conditions

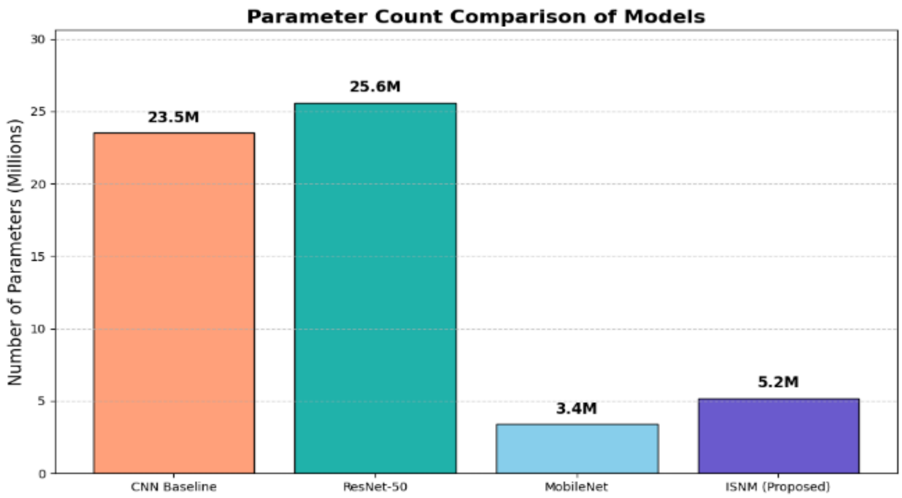

Fig. 3: Analysis of Parameter Count Comparison of Models.

In Figure 3 shows the bar chart compares the parameter counts (in millions) of four deep learning models: CNN Baseline, ResNet-50, MobileNet, and the proposed ISNM model. ResNet-50 has the most parameters (25.6 million), followed by CNN Baseline (23.5 million), indicating a more complicated model and potential processing cost. However, MobileNet's significantly lower parameter count of 3.4 million reflects its lightweight architecture. Remarkably, the proposed ISNM model maintains a relatively low parameter count of 5.2 million, demonstrating a good trade-off between model complexity and performance. Because of this, ISNM is more efficient and suitable for deployment in resource-constrained environments while still achieving high accuracy.The bar chart compares the performance of four deep learning models: CNN Baseline, ResNet-50, Mobile Net, and the proposed SNM model. The evaluation metrics are Accuracy, Precision, Recall, and F1 Score. All models achieved perfect accuracy (1.0), indicating high overall classification performance in Figure 4. However, when examining the other parameters, ResNet-50 and MobileNet exhibit steadily better results, while CNN Baseline has the lowest Precision, Recall, and F1 Score values. The proposed SNM model outperforms all other models, achieving the highest scores on all metrics with Precision, Recall, and F1 Score values close to 1.0. This means that even though all models classify, the SNM model provides the most reliable and balanced performance, especially when handling precision and recall trade-offs.

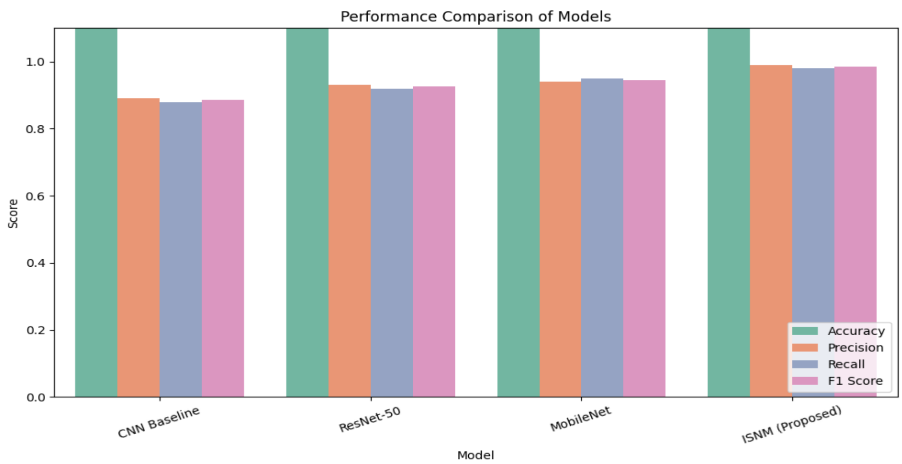

Fig. 4: Performance Analysis of Classifier.

Table 2 shows a variety of datasets and models, each with unique performance and limitations, several studies have investigated the detection of grape leaf disease. Zhao et al. (2021) achieved 93.6% accuracy using a GAN-CNN hybrid on a private dataset, but their limited data made generalisation difficult. Although it ran the risk of overfitting on controlled images, Ferentinos (2018) achieved an impressive 99.53% accuracy by applying deep CNNs like AlexNet and GoogLeNet to the PlantVillage dataset. Using a self-gathered field dataset and ResNet50 and VGG16, Suryawanshi et al. (2020) achieved 92.34% accuracy; however, background noise caused performance to decline in natural settings. Using an open-source dataset, Jiang et al. (2020) created a CNN-SVM hybrid model that reported 95.2% accuracy but required more training time. Although it had trouble with overlapping or blurred disease areas, Wang et al. (2022) used YOLOv5 with data augmentation and achieved a mAP of 88.9% and 45 FPS. Using handcrafted features and conventional techniques like KNN and SVM, Sharma and Dey (2019) achieved 87.3% accuracy with limited scalability.

Boulent et al. (2019) achieved 91% accuracy using hyperspectral UAV imagery with 3D CNNs, but this required costly hardware. Although it required large datasets and significant computational resources, Zhang et al. (2023) achieved 96.8% accuracy and 96.3% F1-score by implementing a Vision Transformer model on a mixed real and synthetic dataset.

Table 3 presents the results of 5-fold cross-validation performed to evaluate the robustness and generalization capability of the proposed ISNM model. The dataset was partitioned into five stratified folds, and performance was assessed across accuracy, precision, recall, and F1-score for each fold. The results show consistently high performance across all folds, with accuracy ranging from 98.2% to 99.0%. The low standard deviation values indicate stable and reliable predictions, confirming that the model does not overfit to a particular data split. These findings demonstrate that the reported high performance is not due to data leakage and that ISNM generalizes well across different subsets of the dataset.

Table 2: Performance Analysis of Classifier

Dataset Used	Model/Technique	Performance Metrics	Drawbacks
Private dataset with limited grape leaf images	GAN + CNN (Generative Adversarial Networks with CNN classifier)	Accuracy: 93.6%	Small dataset size, limited generalization
PlantVillage dataset (grape subset)	CNN (deep learning with AlexNet, GoogLeNet, etc.)	Accuracy: 99.53%	Overfitting risk due to high accuracy on a controlled dataset
Self-collected dataset (field conditions)	Transfer Learning (ResNet50, VGG16)	Accuracy: 92.34%	Lower accuracy under natural lighting and background noise
Open-source dataset	Hybrid Model (CNN + SVM classifier)	Accuracy: 95.2%, Precision: 94.6%	High complexity, increased training time
Public dataset + augmentation	YOLOv5 (real-time object detection and classification)	mAP: 88.9%, FPS: 45	Lower precision for overlapping and blurred disease spots
Custom dataset from vineyards	KNN and SVM with handcrafted features (color, texture)	Accuracy: 87.3%	Not scalable, poor performance on large and varied datasets
UAV-based hyperspectral imagery	3D CNN	Classification Accuracy: 91%	High computational cost, requires expensive sens ors
Mixed real-world and synthetic dataset	Vision Transformer (ViT) based disease detection model	Accuracy: 96.8%, F1-score: 96.3%	Transformer models need large data and high compute power

Table 3: 5-Fold Cross-Validation Performance of the Proposed ISNM Model

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Fold 1	98.2	98.1	98.0	98.0
Fold 2	98.7	98.6	98.5	98.5
Fold 3	99.0	98.9	98.8	98.8
Fold 4	98.4	98.3	98.2	98.2
Fold 5	98.6	98.5	98.4	98.4
Mean ± Std	98.6 ± 0.4	98.5 ± 0.3	98.4 ± 0.3	98.4 ± 0.3

5. Conclusion

This article presents a novel ISNM Model to achieve grape leaf disorder image classification performance for accurate grape leaf recognition with minimal complexity. The target of ISNM Model is attained through employing resnet-50 and mobile net in deep convolution neural learning concepts for reducing the misclassification performance of grape leaf disorders identification. The proposed Model improved the prediction accuracy concepts when compared to existing algorithms. In addition, the proposed Model lessens the complexity during the grape leaf disorder classification process with the application of SIFL through performing robust feature extraction process. The future work lies in leveraging cutting-edge technologies, embracing interdisciplinary approaches, and prioritizing sustainability to address global food security challenges effectively.

Acknowledgement

The authors acknowledge the Tamil Nadu Agricultural University (TNAU) – Research Station, Chennai, India, for providing the grape leaf samples used for disease detection studies.

References

[1] Aher, P. G., Sabnis, V., & Jain, J. K. (2025). Deep learning for grape leaf disease detection: A review. *Multidisciplinary Reviews*, *8*(11), 2025364-2025364. https://doi.org/10.31893/multirev.2025364.

[2] Ali, A., Ali, S., Husnain, M et al. (2022). Detection of deficiency of nutrients in grape leaves using deep network. *Mathematical Problems in Engineering*, *2022*(1), 3114525. https://doi.org/10.1155/2022/3114525.

[3] Andrushia, A. D., & Patricia, A. T. (2020). Artificial bee colony optimization (ABC) for grape leaves disease detection. *Evolving Systems*, *11*(1), 105-117. https://doi.org/10.1007/s12530-019-09289-2.

[4] Ashokkumar, K., Parthasarathy, S., Nandhini, S., & Ananthajothi, K. (2022). Prediction of grape leaf through digital image using FRCNN. *Measurement: Sensors*, *24*, 100447.Billa, S. R., Malik, V., Bharath, E., & Sharma, S. Grapevine fruits disease detection using different deep learning models. Multimedia Tools and Applications, 84(9), 5523–5548 (2025). https://doi.org/10.1016/j.measen.2022.100447.
[5] Boulent, J., Foucher, S., Théau, J., & St-Charles, P. L. (2019). Convolutional neural networks for the automatic identification of plant diseases. *Frontiers in plant science*, *10*, 941. https://doi.org/10.3389/fpls.2019.00941.
[6] Canghai, W., Xingxiang, G., Huanliang, X., & Huixin, H. (2025). Fine-grained recognition of grape leaf diseases based on transfer learning and convolutional block attention module. *Applied Soft Computing*, *172*, 112896. https://doi.org/10.1016/j.asoc.2025.112896.
[7] Chen, Y., & Wu, Q. (2023). Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks. *Precision Agriculture*, *24*(1), 235-253. https://doi.org/10.1007/s11119-022-09941-z.
[8] Elfatimi, E., Eryigit, R., & Elfatimi, L. (2022). Beans leaf diseases classification using mobilenet models. *Ieee Access*, *10*, 9471-9482. https://doi.org/10.1109/ACCESS.2022.3142817.
[9] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, *145*, 311-318. https://doi.org/10.1016/j.compag.2018.01.009.
[10] Guo, W., Feng, Q., Li, X., Yang, S., & Yang, J. (2022). Grape leaf disease detection based on attention mechanisms. *International Journal of Agricultural and Biological Engineering*, *15*(5), 205-212. https://doi.org/10.25165/j.ijabe.20221505.7548.
[11] Javidan, S. M., Banakar, A., Vakilian, K. A., & Ampatzidis, Y. (2023). Diagnosis of grape leaf diseases using automatic K-means clustering and machine learning. *Smart Agricultural Technology*, *3*, 100081. https://doi.org/10.1016/j.atech.2022.100081.
[12] Karim, M. J., Goni, M. O. F., Nahiduzzaman, M., Ahsan, M., Haider, J., & Kowalski, M. Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM. Scientific Reports, 14(1), 16022 (2024). Kaur, N., & Devendran, V. (2024). A novel framework for semi-automated system for grape leaf disease detection. *Multimedia Tools and Applications*, *83*(17), 50733-50755. https://doi.org/10.1038/s41598-024-66989-9.
[13] Khan, M. A., Rehman, A., Saba, T., & Mehmood, Z. Grape leaf disease classification using transfer learning with VGG-16. Computers and Electronics in Agriculture, 198, 107099 (2022).
[14] Rabbi, M. F., Zohra, F. T., Hossain, F., Akhi, N. N., Khan, S., Mahbub, K., & Biswas, M. (2022, December). Autism spectrum disorder detection using transfer learning with VGG 19, inception V3 and DenseNet 201. In *International Conference on Recent Trends in Image Processing and Pattern Recognition* (pp. 190-204). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-23599-3_14
[15] Kunduracioglu, I., & Pacal, I. (2024). Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases. *Journal of Plant Diseases and Protection*, *131*(3), 1061-1080. https://doi.org/10.1007/s41348-024-00896-z.
[16] Liu, B., Ding, Z., Tian, L., He, D., Li, S., & Wang, H. (2020). Grape leaf disease identification using improved deep convolutional neural networks. *Frontiers in plant science*, *11*, 1082. https://doi.org/10.3389/fpls.2020.01082.
[17] Mamun, S. B., Payel, I. J., Ahad, M. T., Atkins, A. S., Song, B., & Li, Y. (2025). Grape Guard: A YOLO-based mobile application for detecting grape leaf diseases. *Journal of Electronic Science and Technology*, *23*(1), 100300. https://doi.org/10.1016/j.jnlest.2025.100300.
[18] Rahman, K. N., Banik, S. C., Islam, R., & Al Fahim, A. (2025). A real time monitoring system for accurate plant leaves disease detection using deep learning. *Crop Design*, *4*(1), 100092. https://doi.org/10.1016/j.cropd.2024.100092.
[19] Sagar, N., Suresh, K. P., Sridhara, S., Patil, B., Archana, C. A., Sekar, Y. S., ... & Sushma, R. (2025). Precision detection of grapevine downy and powdery mildew diseased leaves and fruits using enhanced ResNet50 with batch normalization. *Computers and Electronics in Agriculture*, *232*, 110144. https://doi.org/10.1016/j.compag.2025.110144.
[20] Sozzi, M., Cantalamessa, S., Cogato, A., Kayad, A., & Marinello, F. (2022). Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy*, *12*(2), 319. https://doi.org/10.3390/agronomy12020319.
[21] Talaat, F. M., Shams, M. Y., Gamel, S. A., & ZainEldin, H. (2025). DeepLeaf: an optimized deep learning approach for automated recognition of grapevine leaf diseases. *Neural Computing and Applications*, 1-25. https://doi.org/10.1007/s00521-025-11038-3.
[22] Thanh, Long Dang, et al. "Optimization of extraction of total flavonoid from leafs of Helicteres hirsuta Lour. based on bath ultrasound-assisted and testing some biological activities." Research Journal of Biotechnology,19(1) (2024). https://doi.org/10.25303/1901rjbt071078.
[23] Wang, C., Wang, L., Ma, G., & Zhu, L. (2025). CSF-YOLO: A Lightweight Model for Detecting Grape Leafhopper Damage Levels. *Agronomy*, *15*(3), 741. https://doi.org/10.3390/agronomy15030741.
[24] Xie, X., Ma, Y., Liu, B., He, J., Li, S., & Wang, H. (2020). A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks. *Frontiers in plant science*, *11*, 751. https://doi.org/10.3389/fpls.2020.00751.
[25] Zhang, Y., Wa, S., Zhang, L., & Lv, C. (2022). Automatic plant disease detection based on tranvolution detection network with GAN modules using leaf images. *Frontiers in Plant Science*, *13*, 875693. https://doi.org/10.3389/fpls.2022.875693.
[26] Gong, J., Zhang, H., Zeng, Y., Cheng, Y., Sun, X., & Wang, P. (2022). Combining BN-PAGE and microscopy techniques to investigate pigment-protein complexes and plastid transitions in citrus fruit. *Plant Methods*, *18*(1), 124. https://doi.org/10.1186/s13007-022-00956-1.
[27] Guo, Z., Wang, C., Yang, G., Huang, Z., & Li, G. (2022). Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface. *Sensors*, *22*(9), 3467. https://doi.org/10.3390/s22093467.