

FusionNet-GLD: A Dual-Backbone CNN Model Combining Xception and Inception for Grape Leaf Disease Recognition

Syed Shafia Zainab¹, K.V.Narasimha Reddy², Soudagar Min Haz³, Chatti Navya⁴,
V.Lakshmi⁵, Niharika Duddela⁶,dodda Venkatareddy⁷

shafia2801@gmail.com¹, narasimhareddyneec03@gmail.com², soudagarminhaz@gmail.com³,
chattinavya@gmail.com⁴, lakshmi1682@grietcollege.com⁵, niharikad@gnits.ac.in⁶, doddavenkatareddy@gmail.com⁷

Department of Computer Science and Engineering^{1,2,3,4,5,7},Department of Mechanical Engineering⁶

Narasaraopeta Engineering College^{1,2,3,4,7}, Narasaraopet, Andhra Pradesh, India.

GRIET⁵, G. Narayanamma Institute of Technology & Science (Women)⁶, Hyderabad, Telangana, India.

Abstract—Grapevine leaf diseases like black rot, leaf blight, and esca directly impact vineyard productivity through yield decline and fruit quality deterioration. Early and precise detection will go a long way in efficient management, but conventional manual inspections are time-consuming, labor-intensive, and tend to be unreliable for commercial vineyards. To solve these issues, we introduce FusionNet-GLD, a double-backbone convolutional neural network for the classification of grapevine leaf diseases. It combines Xception and InceptionV3 models to take their complementary advantages. Xception extracts fine-grained patterns through depthwise separable convolutions, whereas InceptionV3 captures features across different scales via small and large convolutional filters. This synergy allows FusionNet-GLD to well capture the intricate visual patterns of affected leaves. The predictive framework was also drilled and assessed on an augmented PlantVillage resource in real vineyard-simulating conditions. The experimental results indicate that FusionNet-GLD beats five baseline models—Random Forest, MobileNetV2, EfficientNetB0, InceptionV3, and Xception—with 99.63% accuracy, 99.45% precision, 99.42% recall, 99.43% F1-score, and 0.99 AUC. These results prove FusionNet-GLD to be efficient, accurate, and scalable. Its lightweight architecture makes it deployable on handheld devices or drones, which will be useful in monitoring vineyards in real time and facilitating precision agriculture by minimizing the use of manual labor and maximizing early disease detection.

Index Terms—Grape leaf disease, FusionNet-GLD, CNN, Xception, InceptionV3

I. INTRODUCTION

I. INTRODUCTION Globally, the grapevine is an important and commercially valuable crop, especially for wine and table grapes. Grape leaf diseases such as Black Rot, Esca, and Leaf Blight can negatively impact yield and fruit quality [1], [2]. In larger vineyards, traditional methods of disease identification are slow and rarely consistent [3]. Therefore, the agricultural sector is increasingly utilizing AI-focused solutions for precision farming. Deep learning, with its ability to recognize patterns and identify diseases based on dataset images, has emerged as a favored option [4], [5]. However, most of these models rely on single-backbone CNNs which

typically yield lower accuracy when attempting to balance speed, computation power, and high accuracy. Prior research has explored various models, including CNNs [6], lightweight CNNs [7], and transformer models [9]. However, to date, few have explored the concept of mixed models that combine the power of multiple frameworks for plant pathology using grapevines as an agricultural example. This study proposes FusionNet-GLD, a dual-backbone deep learning model that utilizes both Xception [10] and InceptionV3 to extract depth-wise and multi-scale spatial features, respectively. We outline four major contributions in the remainder of this study. The architecture of the FusionNet-GLD hybrid CNN model for condition recognition in grape leaves. The contrastive analysis of six algorithms (including a Random Forest model), viz., MobileNetV2, EfficientNetB0, InceptionV3, Xception, and FusionNet-GLD. A detailed performance comparison of models based on performance measures like accuracy, precision, recall, and F1-score, etc.

To confirm FusionNet-GLD, we experimented with an extended PlantVillage dataset, a benchmark for leaf disease classification [6], [11]. Data augmentation enhanced dataset diversity, mimicking real-field scenarios and correcting class imbalance typical of grape leaf disease datasets [5]. All the models—Random Forest, MobileNetV2, EfficientNetB0, InceptionV3, and Xception—were trained and tested under similar conditions. The performance was evaluated by accuracy, precision, recall, F1-score, and AUC, allowing for rigorous comparison [3], [4], [10].

II. LITERATURE REVIEW

Recently, lightweight detection models have become prominent for real-time implementation. For example, the YOLOv8-ACCW model features a lean architecture that includes AK-Conv, Coordinate Attention (CA), and CARAFE modules. It reports an F1 score of 92.4% and a mAP50 of 92.8% while maintaining a model size of less than 2.8MB [7]. Similarly, the GCS-YOLO model combines GhostNet modules

with CBAM attention to achieve a mAP@0.5 of 96.2% with only 1.63 million parameters. This approach performs well under variable lighting conditions and is suitable for edge deployment [8].

Dual-path architectures can effectively extract local and global features, thus making them very credible for plant disease classification. GrapeLeafNet is a good example, which combines Inception-ResNet, CBAM, and a Shuffle-Transformer in an effort to enhance feature representation. Using this architecture, the model attained 99.56% accuracy on the PlantVillage dataset [9]. To address the challenges of real-time deployment on edge devices, Karim et al. employed a tuned MobileNetV3 Large with custom dense layers and Grad-CAM visualizations. On an Nvidia Jetson Nano, this approach achieved over 99.4% accuracy on both training and test sets [12].

The presence of well-structured datasets is important for building trustable models. For instance, the NGLD dataset contains 2,726 Indian vineyard annotated images employed in training models that identify downy mildew and bacterial leaf spot [11].









Additionally, some studies have compared the performance of CNN-based models to transformer models. One such report indicated that, upon proper fine-tuning, a Vision Transformer (Swinv2-Base) reached 100% validity on the PlantVillage and grape variety collection [13]. Addressing model interpretability, the DeepLeaf approach uses fuzzy-optimized CNNs and logistic regression for downsampling, achieving 99.7% accuracy while managing class imbalance and feature redundancy [14]. Research by Mangaoang et al. highlights the effectiveness of models like MobileNetV2 and EfficientNet. Their trials demonstrated that a fine-tuned EfficientNet could achieve 100% accuracy, underscoring its promising role in precision agriculture [15]. Although traditional machine learning techniques do not match the flexibility of deep learning, they are still used for grape leaf classification, offering interpretability and computational efficiency for certain food quality applications [5]. Furthermore, Utam demonstrated the effectiveness of CNNs like VGG16 and MobileNet for automating disease detection, achieving accuracies up to 97%, which can expedite early diagnosis and enhance vineyard management [6].

III. METHODOLOGY

A. Overview of the Dataset

We use the fully publicly available, PlantVillage grape leaf dataset, available publicly through Kaggle and one of the augmented versions published in recent research papers [11]. This dataset consists of high-resolution images, represented in five categories (i.e., plant diseases, healthy samples) including Black rot, Esca (also known as Black measles), Leaf Blight (also known as Isariopsis leaf spot), and Healthy samples. The dataset has been augmented using a variety of transformations to ensure diverse samples in all classes. Classes were balanced to include 1000 images in each class.

TABLE I: Categories of Grape Leaf Diseases with Example Images

Class	Sample Image 1	Sample Image 2
Black rot		
Esca		
Leaf blight		
Healthy		

B. Data Preprocessing

To maintain uniformity and enhance model efficiency, multiple preprocessing operations were carried out:

- **Image Resizing:** whole pictures were cropped to a uniform size that is 224×224 pixels for maintaining uniform input size.
- **Pixel Scaling:** RGB intensity metrics have been adjusted to limits $[0, 1]$, which improved convergence rates during training.
- **Data Augmentation:** Real-time flips, random rotations, zooming, and contrast changes were applied in real time using the ImageDataGenerator module to enhance dataset diversity and prevent overfitting.
- **Dataset Partitioning:** The dataset was partitioned into three sets, with **70% training**, **15% validation**, and **15% testing**.

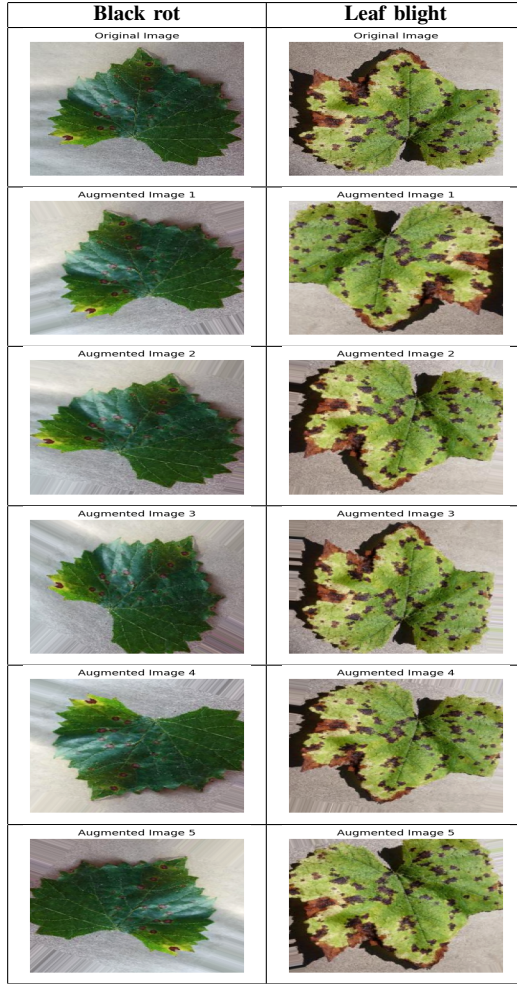
C. Architectures

The following six models were implemented and compared:

- **Random Forest (RF):** A classical machine learning baseline using hand-crafted features.
- **MobileNetV2:** A lightweight convolutional neural network (CNN) architecture with inverted residual blocks.
- **EfficientNetB0:** A balanced and scalable CNN architecture, optimized using neural architecture search.
- **InceptionV3:** Leverages multi-scale convolution operations using Google's inception modules.

- **Xception:** Similar to InceptionV3 but uses depth-wise separable convolutions instead of standard convolutions.
- **FusionNet-GLD (Proposed):** A dual-branch fusion model that combines Xception and InceptionV3 architectures in parallel to enhance feature extraction and classification performance.

TABLE II: Images of Dark rot lesions and Vine leaf infection prior to and following augmentation



D. FusionNet-GLD: Architecture

This model consists of two parallel branches:

- **Xception Branch:** Employed a depthwise separable convolutional neural networks to extract strong high-level discriminative features.
- **InceptionV3 Branch:** Employs Inception modules to extract features from multiple receptive fields, providing the model with the ability to extract fine-grained features while capturing more context on a larger scale.
- **Fusion Layer:** Feature maps between the dual divisions are combined and then passed through:
 - spatial global averaging layer
 - linear layer with ReLU activation,

- and predicted class probabilities for multi-class classification.

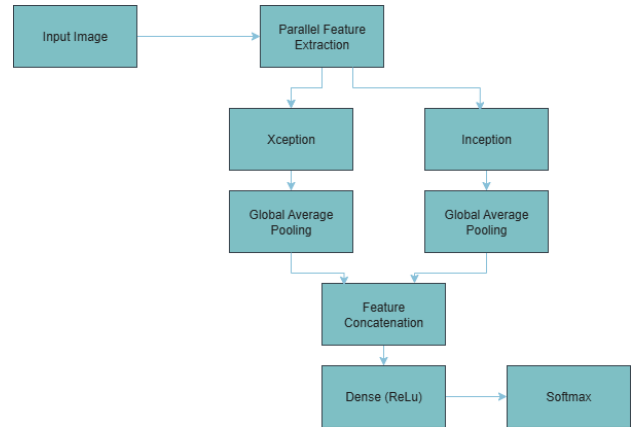


Fig. 1: Architecture of the proposed FusionNet-GLD model combining Xception and InceptionV3 in a dual-branch structure.

Fig 1 provides a flowchart of the newly proposed framework, FusionNet-GLD. It is a dual-branch framework with both Xception and Inception networks applying a parallel feature extraction approach. Utilizing a dual-branch framework provides benefits from both unique aspects of each architecture (1) the depth-wise separable convolutions are utilized to create deep and efficient representations of the target and (2) there are several convolutional layers within each branch which are used to extract both global and local characteristics from the images of diseased leaves. After feature extraction, global mean sub-sampling is applied to reduce redundancy features, and mitigate overfitting in all of the branches in the proposed framework. After sub-sampling, the two branches are concatenated into a single vector, then passed into the Dense layer with a ReLU activation function that allows the model to learn deep and complex representations of non-linear data features, as well as capabilities to capture representations of the data features. On the final layer of the model, the SoftMax classifier generates a probabilistic distribution across the target disease classification classes with a reliable way to classify. This hybrid feature fusion approach achieves increasing discriminative power of the model and provides high classification performance for grape leaf disease classification.

E. Training Configuration

The training of the FusionNet-GLD model was conducted with optimal hyperparameters, learning methods, and regularization methods to achieve good convergence and generalization. The training conditions are as follows:

- **Optimizer:** An Adam optimizer was chosen because of its adaptive learning rates for each weight and rapid convergence with the new gradient update, with a gradient step size of 0.0001. The optimizer computes exponentially decayed averages of past gradients and their squares

(i.e., moving averages) and thus retains the benefits of RMSProp and momentum as well as shown in Eq. (1).

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \\ \hat{v}_t &= \frac{v_t}{1 - \beta_2^t}, \\ \theta_{t+1} &= \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad (1)$$

- **Loss Function:** The problem tackled multi-class classification, the Categorical Cross-Entropy loss was utilized. This function measures the error between the predicted class probabilities and the truth labels, mathematically formulated as

$$\mathcal{L} = - \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (2)$$

- **Learning iterations:** To mitigate overfitting, model training was limited to a maximum of 50 epochs, with early stopping employed as a regularization strategy. Training was stopped if, after five consecutive epochs, no improvement in validation loss was observed.
- **Batch Size:** A mini-batch size of 32 was utilized to reach an optimal trade-off between computational expense and training stability.
- **Learning Rate Schedule:** To enable stable convergence and prevent the optimizer from becoming trapped in local minima, the ReduceLROnPlateau callback was utilized. This method dynamically reduces the learning rate by a factor of 0.1 whenever the validation loss does not improve over three to five consecutive epochs.
- **Validation Partitioning:** To test the model's generalization capacity during training, 20% of the training set was held aside for validation.
- **Initialization of Weight:** The He Normal Initialization technique, sometimes referred to as Kaiming initialization, was employed to initialize weights as follows:

$$W \sim \mathcal{N}\left(0, \sqrt{\frac{2}{n_{\text{in}}}}\right) \quad (3)$$

- **Model Checkpointing:** The model's best weights, in terms of validation accuracy, were saved during training to make sure that the final evaluation was based on the best set of parameters.
- **Data Augmentation and Shuffling:** To avoid bias in the learning order of the data, the input data was shuffled before each epoch. Data augmentation processes, such as random rotation, flipping both row and column, zooming, and brightness variation, were also done to improve robustness and reduce overfitting.

IV. RESULTS

In order to measure the performance of the developed models, a set of experiments was conducted on the dataset of grape

leaf disease. The measurement was based on important performance metrics, i.e., Accuracy, Precision, Recall, F1-Score, and the Area Under the ROC Curve (AUC). Each of these collectively measures the overall classification capability of a model capability to differentiate the good and damaged leaf.

A. Evaluation Metrics

The key classification metrics used in this study are formally represented in the following manner, where TP, TN, FP, and FN denote actual positives, actual negatives, false positives, and false negatives, accordingly:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{Precision} &= \frac{TP}{TP + FP}, \\ \text{Recall} &= \frac{TP}{TP + FN}, \\ \text{F1 Score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \\ \text{AUC-ROC} &= \int_0^1 \text{TPR}(x) d(\text{FPR}(x)) \end{aligned} \quad (4)$$

B. Comparison of Model Performance

TABLE III: summarizes the performance of all six implemented models on the test set. FusionNet-GLD achieved the highest scores across all metrics, demonstrating superior classification capability.

Model	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	87.65%	87.01%	86.45%	86.73%	0.89
MobileNetV2	95.82%	95.60%	95.20%	95.40%	0.97
EfficientNetB0	96.88%	96.55%	96.33%	96.44%	0.98
Inception	97.34%	97.10%	96.80%	96.95%	0.98
Xception	98.24%	98.12%	97.94%	98.03%	0.98
FusionNet-GLD	99.63%	99.45%	99.42%	99.43%	0.99

The superior performance of FusionNet-GLD is attributed to its dual-feature extraction capabilities, effectively leveraging deep and multi-scale features to capture subtle disease patterns.

C. Discussion

Compared to GrapeLeafNet [9], a more complex model integrating transformers and Inception-ResNet which achieved 99.56% accuracy, the proposed simpler FusionNet-GLD model achieves a slightly higher accuracy of 99.63%. Additionally, this classification-based approach differs from YOLOv8 [7], [8], which is geared towards object detection rather than image-level classification.

D. Training and Validation

The proposed method was trained and validated over 10 augmented iterations of the grape leaf dataset. Fig 2 and Fig 3 show a steady increase in accuracy and a decrease in validation loss, indicating good model convergence and generalization.

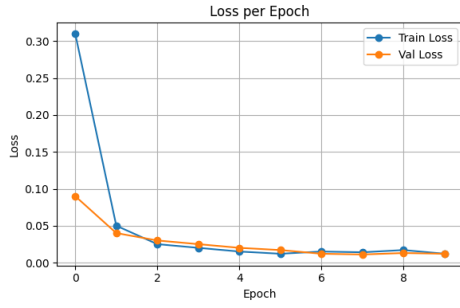


Fig. 2: Training and validation loss per epoch.

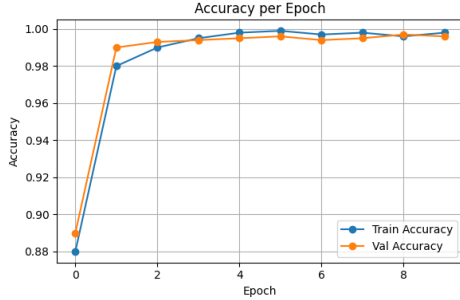


Fig. 3: Training and validation accuracy per epoch.

E. Confusion Matrix Analysis

Fig 4 displays the confusion matrix on the validation set, indicating highly accurate classification across all grape leaf disease categories. Most classes exhibit near-perfect true positive rates with minimal misclassification.

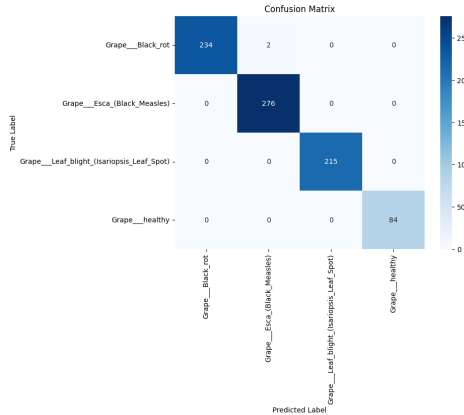


Fig. 4: Confusion matrix of the FusionNet-GLD model on the validation set.

Specifically, the model correctly classified 234 out of 236 samples for **Grape_Black rot** with 2 misclassifications; it perfectly classified all 276 samples of **Grape_Esca (Black Measles)**; and all 215 samples of **Grape_Leaf blight (Isariopsis Leaf Spot)**. Additionally, it achieved flawless classification of all 84 **Grape_Healthy** samples.

This outstanding performance across classes validates the model's efficacy in learning stable, disease-specific features

and generalizing well to unseen data, supporting the reliability of FusionNet-GLD for grape leaf disease identification.

F. Model Comparison Visualization

Fig 5 visualizes the accuracy comparison of FusionNet-GLD against baseline models, highlighting its superiority in prediction accuracy.

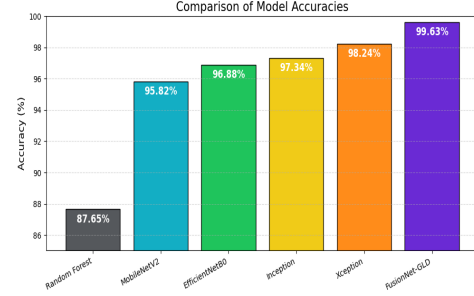


Fig. 5: Model accuracy comparison across different architectures.

G. Comparison with Existing Works

The following table presents a comparison of various prior detection of Grapevine disorder. It appears that your proposed FusionNet-GLD model using both Xception and, Inception architectures has superior accuracy. Therefore, your proposed dual-backbone fusion method is beneficial in classification tasks for agricultural applications. The decreased false positives and false negatives is likely improvement in classification performance due to the complementary strengths of the Xception and Inception architectures where Xception is effective in capturing depth wise spatial features while Inception is effective in identifying patterns that exist at multiple scales. In addition to improving the classification performance, the model can have a noticeably lesser false positive and false negative when classifying disease in real-world applications.

TABLE IV: The performance comparison of FusionNetGLD to other models.

S.No	Source	Methodology	Accuracy
1	Uttam, A. K. [6]	CNN with pre-processing	96.42%
2	Karthik, R. et al. [9]	Inception-ResNet + Shuffle Transformer	99.56%
3	Khirade & Patil [3]	Basic CNN with image processing	92.50%
4	Sonar & Wankhade [5]	ML Classifiers (SVM, RF, KNN)	92.60%
5	Proposed Work	Xception + Inception (Fusion CNN)	99.63%

V. CONCLUSION

In this research study, the structural model employed is a double-backbone deep learning model called FusionNet-GLD, which combines Xception and InceptionV3 architectures to

accurately classify grapevine leaf diseases. The model's performance was validated by comparison with existing models, demonstrating superior effectiveness.

This study contributes significantly to smart agriculture by providing a useful tool for early disease detection, enabling farmers to take timely action to minimize crop damage.

A. Future Scope

Potential Extensions could prioritize deploying classifier in physical world on smartphones or edge computing systems to assist farmers directly in the field. Integration with Internet of Things (IoT) systems, drones, or satellite technology could facilitate vineyard monitoring at larger scales.

Further research could explore using multimodal data, including images of grapevines combined with soil condition information, to enhance diagnostic accuracy. Employing transfer learning or self-supervised learning techniques may also enable adapting the model for disease detection in other crops or varying environmental conditions.

Additionally, lightweight model compression and optimization methods such as pruning or quantization could be investigated to reduce computational costs, making the model suitable for deployment on low-resource devices and immediate applications in hard-to-reach regions with poor connectivity.

If these approaches were adopted in addition to building upon the existing model, FusionNet-GLD could evolve into a complete decision-support system for grapevine diseases, functioning as a strong and scalable AI solution for crop health management. This improvement would be a substantial step toward food security, reducing dependence on manual checking for diseases, and supporting future sustainable agriculture driven by AI.

REFERENCES

- [1] A. Seccia, F. G. Santeramo, and G. Nardone, "Trade competitiveness in table grapes: a global view," *Outlook on Agriculture*, vol. 44, no. 2, pp. 127–134, 2015.
- [2] Wikipedia contributors, "Black rot (grape disease)," *Wikipedia, The Free Encyclopedia*, May 30, 2025.
- [3] S. D. Khirade and A. B. Patil, "Plant disease detection using image processing," in *Proc. Int. Conf. Computing Communication Control and Automation*, pp. 768–771, 2015.
- [4] V. Kanabur, S. S. Harakannanavar, V. I. Purnikmath, P. Hullole, and D. Torse, "Detection of leaf disease using hybrid feature extraction techniques and CNN classifier," in *Proc. Int. Conf. Computational Vision and Bio Inspired Computing*, Cham: Springer, pp. 1213–1220, 2019.
- [5] P. S. Sonar and N. R. Wankhade, "Grape leaf disease identification using machine learning techniques," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 4, no. 7, 2022.
- [6] A. K. Uttam, "Grape leaf disease prediction using deep learning," in *Proc. Int. Conf. Applied Artificial Intelligence and Computing (ICAAIC)*, pp. 369–373, May 2022.
- [7] Z. Chen, J. Feng, Z. Yang, Y. Wang, and M. Ren, "YOLOv8-ACCW: Lightweight grape leaf disease detection method based on improved YOLOv8," *IEEE Access*, 2024.
- [8] Q. Hu and Y. Zhang, "GCS-YOLO: A Lightweight Detection Algorithm for Grape Leaf Diseases Based on Improved YOLOv8," *Applied Sciences*, vol. 15, no. 7, p. 3910, 2025.
- [9] R. Karthik et al., "GrapeLeafNet: A dual-track feature fusion network with inception-ResNet and shuffle-transformer for accurate grape leaf disease identification," *IEEE Access*, vol. 12, pp. 19612–19624, 2024.
- [10] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1251–1258, 2017.
- [11] M. Dharrao et al., "Grapes leaf disease dataset for precision agriculture," *Data in Brief*, 2025.
- [12] M. J. Karim et al., "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight CNN architecture and Grad-CAM," *Scientific Reports*, vol. 14, no. 1, p. 16022, 2024.
- [13] I. Kunduracioglu and I. Pacal, "Advancements in deep learning for accurate classification of grape leaves and diagnosis of grape diseases," *J. Plant Dis. Prot.*, vol. 131, no. 3, pp. 1061–1080, 2024.
- [14] F. M. Talaat, M. Y. Shams, S. A. Gamel, and H. ZainEldin, "DeepLeaf: an optimized deep learning approach for automated recognition of grapevine leaf diseases," *Neural Comput. Appl.*, pp. 1–25, 2025.
- [15] E. F. Mangaoang, T. D. Palaoag, and J. S. Ingosan, "Analysis of deep learning algorithms for grape leaf disease detection," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 33s, pp. 336–344, Apr. 2025.
- [16] P. M. S. Lakshmi, K. LakshmiNadh, K. N. Reddy, and D. V. Reddy, "Ensemble-based transfer learning for multi-class plant disease detection using VGG16, ResNet50, and Xception models," in *Proc. 2024 Int. Conf. IoT Based Control Networks and Intelligent Systems (ICICNIS)*, pp. 1103–1110, Dec. 2024.
- [17] A. V. Kumar, S. V. N. Sreenivasu, and K. V. N. Reddy, "ResNet-CNN model for plant disease classification for e-agriculture applications," in *Proc. 2024 Int. Conf. Intelligent Algorithms for Computational Intelligence Systems (IACIS)*, pp. 1–7, Aug. 2024.
- [18] K. V. Narasimha Reddy and E. Brahma Reddy, "Crop yield prediction based on weather and soil parameters using regression tree model," in *Proc. Int. Conf. Communication, Devices and Computing*, Singapore: Springer Nature Singapore, pp. 1–10, Mar. 2023.
- [19] C. Venkatachalam, P. Shah, K. R. KM, Y. Kumaran, and A. Roy, "Advanced Grape Leaf Disease Diagnosis Using EfficientNetV2L with Data Augmentation and Grad-CAM Visualization in Precision Agriculture," *Procedia Computer Science*, vol. 260, pp. 332–340, 2025.
- [20] Z. Fu, L. Yin, C. Cui, and Y. Wang, "A lightweight MHD-DETR model for detecting grape leaf diseases," *Frontiers in Plant Science*, vol. 15, p. 1499911, 2024.
- [21] N. Sagar, K. P. Suresh, S. Sridhara, B. Patil, C. A. Archana, Y. S. Sekar, et al., "Precision detection of grapevine downy and powdery mildew diseased leaves and fruits using enhanced ResNet50 with batch normalization," *Computers and Electronics in Agriculture*, vol. 232, p. 110144, 2025.