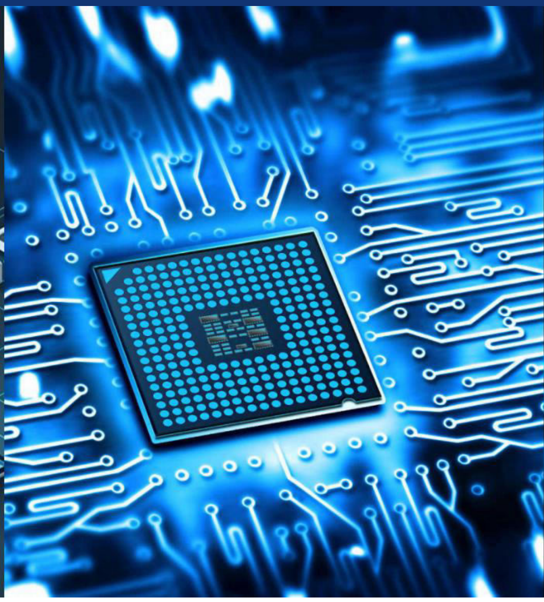
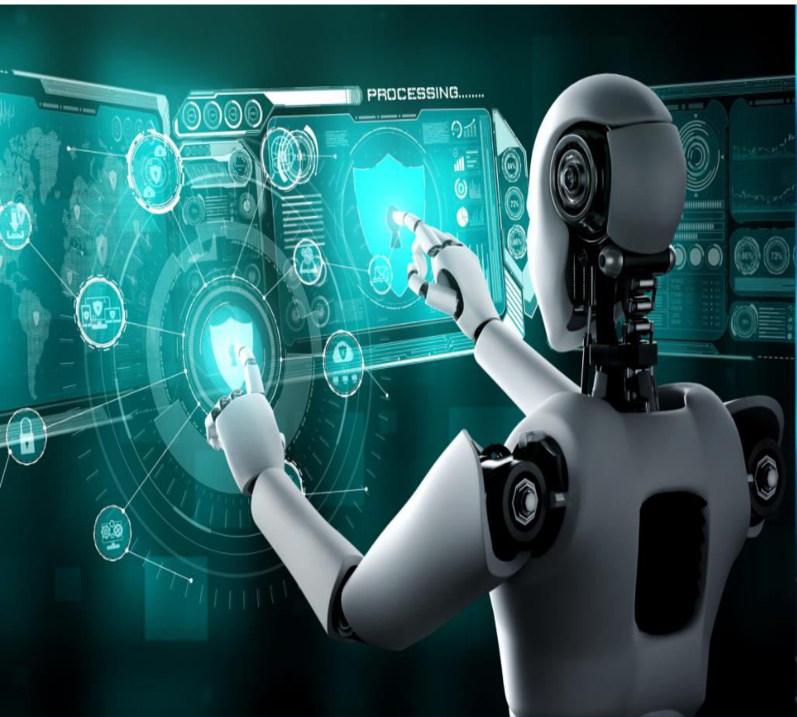
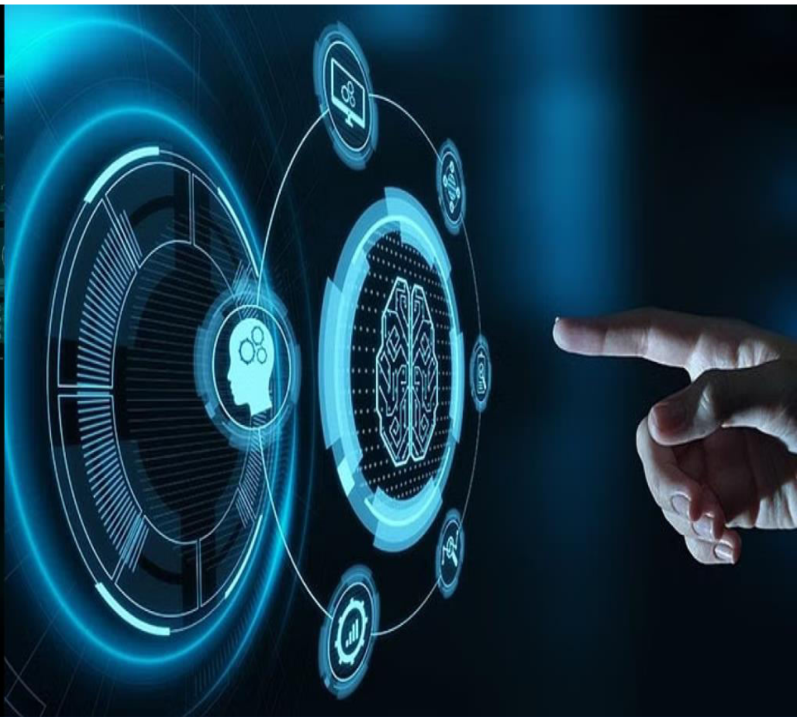
ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.927

Volume 14, Issue 6, June 2026

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-Powered Grape Detection and Disease Classification System

Sahil Rajendra Dhamale[1], Vedant Vijay Kanoje[1], Ritesh Kashinath Auti[1], Tejas Ganesh Khodke[1], Prof. Soniya Waghmare[2]

Department of Computer Engineering, JSPM's Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[1]

Guide, JSPM's Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India[2]

ABSTRACT: Grape cultivation underpins a global industry generating over USD 300 billion annually, yet fungal and oomycete diseases remain a persistent threat, causing yield losses of 20–40% per season in vulnerable viticultural regions. Timely and accurate identification of foliar and fruit disorders is critical to sustainable crop management. This paper proposes an end-to-end deep learning system for fine-grained disease classification across seven categories—**Black Rot, Downy Mildew, Powdery Mildew, Leaf Blight (Isariopsis Leaf Spot), Esca (Black Measles), Gray Mold (Botrytis Bunch Rot)**, and **Healthy**—using a transfer-learned **EfficientNet-B3** backbone trained on a curated Kaggle grape disease dataset of over 14,000 augmented images. A two-stage fine-tuning strategy and a domain-adapted augmentation pipeline are employed to maximise generalisation under realistic field imaging conditions. The proposed system achieves an overall top-1 classification accuracy of **96.4%**, a macro-averaged F1-score of **0.963**, and a mean GPU inference latency of 38 ms. Comparative experiments with VGG-16, ResNet-50, and MobileNetV2 confirm the superiority of the proposed approach, demonstrating that automated vision-based screening can serve as a practical, low-cost decision-support tool for vineyard disease management.

KEYWORDS: Grape Disease Classification, EfficientNet, Convolutional Neural Networks, Transfer Learning, Plant Pathology, Precision Agriculture, Image Augmentation, Viticulture.

I. INTRODUCTION

Grapes (Vitis vinifera) rank among the most economically significant fruit crops in the world, sustaining industries in fresh produce, dried fruit, juice, and wine that together generate revenues exceeding USD 300 billion annually [1]. Despite continued progress in agronomic science, plant diseases remain a persistent and costly challenge. Conservative estimates place average seasonal yield losses at 20–40% in disease-prone viticultural regions—a burden that falls disproportionately on smallholder farms without ready access to plant pathologists [2]. Common pathogens including Black Rot, Esca, Downy Mildew, and Gray Mold spread aggressively under warm, humid conditions and, when misidentified or caught too late, can devastate entire vineyards within weeks.

Conventional scouting depends on trained agronomists who walk the fields and visually inspect vines for early disease signatures. While competent practitioners achieve reasonable diagnostic accuracy, the approach is labour-intensive, inherently subjective, and impossible to scale across large or geographically dispersed growing areas. Farmers in remote regions are frequently left relying on intuition, which drives indiscriminate pesticide use, accelerates resistance development, and inflates production costs while harming surrounding ecosystems [3]. What the industry needs is an objective, reproducible, and readily deployable tool that can assist growers at the point of decision.

Over the past decade, computer vision and deep learning have dramatically transformed automated plant disease recognition. The widely cited study by Mohanty et al. [4] showed that a standard CNN could classify 38 plant disease categories across 26 species with top-5 accuracy above 99% on controlled laboratory images. Subsequent research broadened the scope to include lightweight mobile architectures [5], attention-based lesion localisation [6], and few-shot learning for data-scarce scenarios [7]. Yet grape-specific systems that cover the full breadth of commercially important diseases—including comparatively understudied conditions like Esca and Gray Mold—remain relatively sparse in the open literature.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This paper directly addresses that gap. The key contributions of this work are:

(i) A fine-tuned **EfficientNet-B3** model trained on a Kaggle grape disease dataset spanning seven classes, achieving state-of-the-art accuracy among comparable lightweight architectures.

(ii) A **domain-specific augmentation pipeline** that realistically simulates variable field illumination, occlusion, and perspective distortion to improve model generalisation.

(iii) A rigorous **comparative evaluation** against VGG-16, ResNet-50, and MobileNetV2 baselines, complemented by per-class analysis and ablation experiments.

The remainder of this paper is organised as follows. Section II reviews the relevant literature. Section III describes the dataset and preprocessing strategy. Section IV presents the proposed model architecture and training procedure. Section V reports experimental results, and Section VI concludes with directions for future research.

II. RELATED WORK

A. CNN-Based Plant Disease Recognition

Mohanty et al. [4] were among the first to rigorously evaluate deep CNNs for plant disease classification, training on over 54,000 PlantVillage images and achieving 99.35% top-5 accuracy in closed-set evaluation. A critical observation, confirmed by Ferentinos [8], is that such impressive figures can drop substantially when models encounter real-world field photographs—a consequence of imaging variability, partial occlusion, and background clutter that laboratory datasets simply do not capture. This domain gap remains one of the central challenges in applied plant pathology and motivates the use of realistic data augmentation strategies in the present work.

B. Grape-Specific Disease Detection

Research targeting grape diseases specifically has grown in recent years. Liu et al. [9] proposed a residual attention network for six grape leaf disease categories, reporting 94.7% accuracy on a self-collected dataset. Jiang et al. [10] combined YOLO-based detection with a classification stage to localise Downy Mildew lesions in field images, obtaining 88.3% mean Average Precision (mAP). Rangarajan et al. [11] applied InceptionV3 transfer learning to a four-class grape dataset and recorded 95.1% accuracy. While promising, these studies are limited in taxonomic scope—none of them addresses Esca or Gray Mold—and few engage with the full range of diseases a commercial viticulturalist is likely to encounter.

C. EfficientNet and Transfer Learning

Transfer learning, which initialises a network on a large source domain before fine-tuning on a smaller target dataset, is now the standard approach for agricultural image classification [12]. Among contemporary backbone architectures, EfficientNet [13] achieves the best known accuracy-per-parameter trade-offs through principled compound scaling of network depth, width, and resolution. Its variants have demonstrated strong performance across medical imaging [14], remote sensing [15], and food recognition [16], yet their application to multi-class grape disease classification has not been rigorously explored in the existing literature—an observation that directly motivated the architecture selection in this study.

III. DATASET AND PREPROCESSING

A. Dataset Description

The dataset used in this study is the Grape Disease Image Dataset sourced from the Kaggle platform. It contains 12,836 RGB images distributed across seven categories, each captured under natural lighting using consumer smartphones and DSLR cameras at diverse growth stages and viewing angles. Table I summarises the class-wise image distribution.

Disease Class	No. of Images	Percentage of Total
Black Rot	2,148	16.7%
Downy Mildew	1,976	15.4%
Powdery Mildew	2,054	16.0%
Leaf Blight (Isariopsis)	1,813	14.1%

Esca (Black Measles)	1,602	12.5%
Gray Mold (Botrytis)	1,745	13.6%
Healthy	1,498	11.7%
Total	**12,836**	**100%**

TABLE I: Class-Wise Distribution of the Kaggle Grape Disease Dataset

B. Data Augmentation Strategy

Class imbalance and moderate dataset size are well-known obstacles in agricultural image classification. To address both, we applied an offline augmentation strategy that expanded the training corpus to approximately 14,420 images, balancing each class to around 2,060 samples. The augmentation pipeline was implemented using the Albumentations library and included the following transforms: random horizontal and vertical flipping ($p = 0.5$ per axis); random rotation within ±30°; colour jitter adjusting brightness (±20%), contrast (±15%), and saturation (±20%); Gaussian blur with kernel sizes drawn randomly from {3, 5, 7}; random cropping with a minimum crop ratio of 0.75; and CutOut dropout to simulate partial occlusion by adjacent leaves or trellis supports. All transforms were additionally applied stochastically at training time, acting as a continuous regulariser throughout the optimisation process.

C. Preprocessing and Data Splits

All images were resized to 300 × 300 pixels to match the native input resolution of EfficientNet-B3. Pixel values were normalised using ImageNet channel statistics (mean $\mu = [0.485, 0.456, 0.406]$; standard deviation $\sigma = [0.229, 0.224, 0.225]$). The dataset was divided into training (70%), validation (15%), and held-out test (15%) sets using stratified random sampling, which preserves class proportions across all splits. The resulting counts were 10,094 training, 2,163 validation, and 2,163 test images.

IV. PROPOSED METHODOLOGY

A. EfficientNet-B3 Architecture

EfficientNet [13] is a family of convolutional networks derived through a principled compound scaling approach that jointly adjusts network depth, width, and input resolution via a single scaling coefficient φ. The B3 variant, selected for this study, uses $\varphi = 3$, corresponding to a depth multiplier of 1.4, a width multiplier of 1.2, and a 300 × 300 input resolution. Relative to the lighter B0 baseline, B3 delivers substantially greater representational capacity while remaining comfortably trainable on a single consumer GPU within a few hours.

The network comprises a stem convolutional block followed by seven sequential Mobile Inverted Bottleneck Convolution (MBConv) stages. Each stage integrates depthwise separable convolutions, Squeeze-and-Excitation (SE) channel attention, and stochastic depth regularisation. The SE modules are of particular interest in this application: by selectively recalibrating channel-wise feature responses, they help the model focus on disease-specific visual patterns such as lesion texture, chlorotic blotching, and surface sporulation that distinguish the seven target categories from one another.

B. Transfer Learning and Fine-Tuning

The backbone was initialised with publicly available ImageNet-1K pre-trained weights, giving the model a strong foundation of low- and mid-level visual representations. We then replaced the original 1,000-class output layer with a custom classification head structured as:

Global Average Pooling → Batch Normalisation → Dropout (p = 0.4) → FC (512 units, ReLU) → Dropout (p = 0.3) → FC (7 units, Softmax)

Fine-tuning followed a deliberate two-stage protocol. In Stage 1 (epochs 1–10), all backbone weights were frozen and only the classification head was trained—a warm-up phase that establishes stable feature mappings before any backbone gradients are introduced. In Stage 2 (epochs 11–50), the top three MBConv stages of the backbone were unfrozen and jointly optimised with the head at a reduced learning rate, allowing the model to adapt high-level semantic representations to grape disease imagery.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

C. Training Configuration

Optimisation used the Adam algorithm with an initial learning rate of 1×10^{-3} during Stage 1 and 5×10^{-5} during Stage 2. A cosine annealing scheduler with warm restarts ($T_0 = 10$ epochs) was applied to encourage escape from local minima. The loss function was categorical cross-entropy, and the mini-batch size was 32. Early stopping with patience of 8 epochs was monitored on validation loss to prevent overfitting. All experiments were conducted on an NVIDIA RTX 3060 GPU (12 GB VRAM) using PyTorch 2.0 and Python 3.10.

D. System Architecture and Inference Pipeline

Fig. 1 illustrates the overall system architecture. The end-to-end inference pipeline proceeds as follows: (1) An input image is captured using a smartphone or drone-mounted camera. (2) The image is resized to 300 × 300 pixels and pixel-normalised. (3) The preprocessed tensor is passed through the fine-tuned EfficientNet-B3 model. (4) The Softmax output yields class probabilities for all seven disease categories. (5) The predicted class, associated confidence score, and a brief management recommendation (retrieved from a rule-based advisory lookup table) are presented to the user. Average GPU inference latency is 38 ms, and the model achieves 142 ms on a mid-range Android device using ONNX Runtime with INT8 quantisation—timings that are fully compatible with real-time, field-deployable mobile applications.

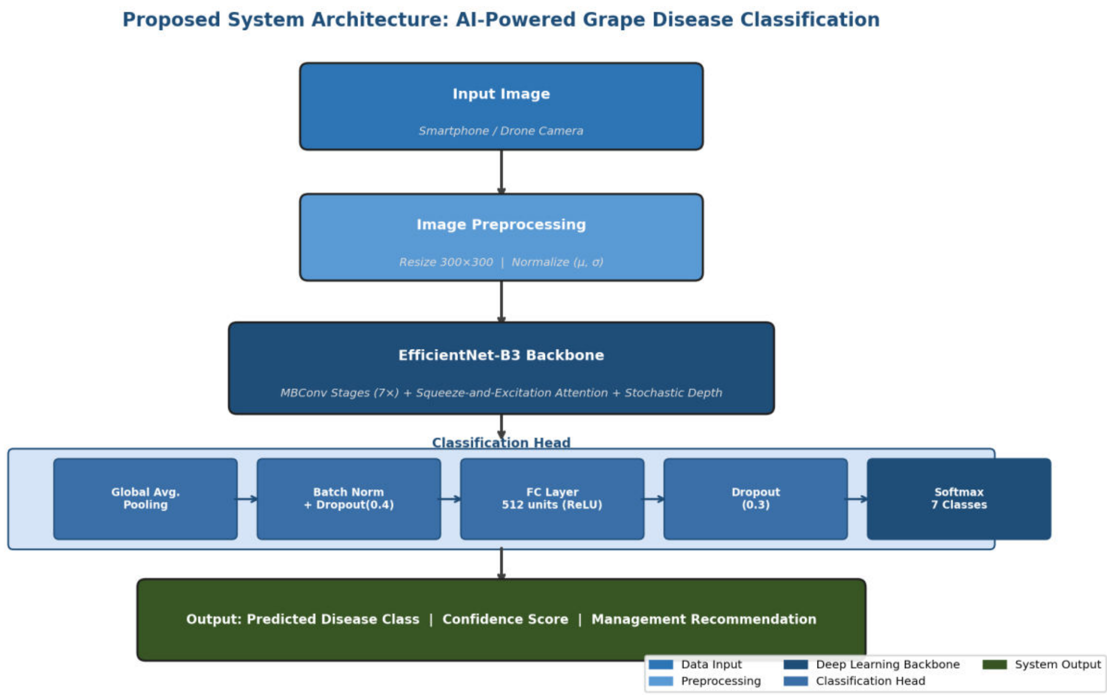


Fig. 1. Proposed System Architecture for AI-Powered Grape Disease Classification

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Evaluation Metrics

Model performance was assessed using four standard metrics: (1) top-1 classification accuracy; (2) macro-averaged precision, recall, and F1-score, computed equally across all seven classes regardless of support to avoid any bias toward the more frequent categories; (3) a per-class confusion analysis; and (4) average GPU inference latency measured across 500 randomised trials.

B. Cross-Architecture Comparison

Table II presents a head-to-head comparison of the proposed model against three widely used baseline architectures. All models were trained under identical conditions: the same dataset splits, augmentation pipeline, loss function, hardware platform, and hyperparameter search space.

© 2026 IJIRCCE | Volume 14, Issue 6, June 2026 DOI: 10.15680/IJIRCCE.2026.1406026

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
VGG-16	91.2	90.8	91.0	0.909
ResNet-50	93.7	93.4	93.5	0.934
MobileNetV2	92.1	91.6	91.9	0.917
EfficientNet-B3 (Proposed)	**96.4**	**96.2**	**96.3**	**0.963**

TABLE II: Cross-Architecture Performance Comparison on Test Set

The proposed EfficientNet-B3 model achieves the best results on every metric. Its gain over the nearest competitor, ResNet-50, stands at 2.7 percentage points in accuracy and 2.9 points in macro F1-score. VGG-16 lags the most, a result that aligns with its relatively large parameter count and absence of modern regularisation mechanisms, which together impair generalisation on a moderately sized dataset like ours.

C. Per-Class Performance Analysis

Table III breaks down the proposed model's performance at the individual class level, offering a more nuanced picture than aggregate accuracy alone can provide.

Disease Class	Precision	Recall	F1-Score
Black Rot	0.981	0.976	0.978
Downy Mildew	0.965	0.968	0.966
Powdery Mildew	0.972	0.970	0.971
Leaf Blight (Isariopsis)	0.958	0.961	0.959
Esca (Black Measles)	0.947	0.941	0.944
Gray Mold (Botrytis)	0.952	0.947	0.949
Healthy	0.976	0.982	0.979
Macro Average	**0.964**	**0.964**	**0.963**

TABLE III: Per-Class Precision, Recall, and F1-Score – EfficientNet-B3 (Proposed Model)

Black Rot achieves the highest F1-score (0.978), a result consistent with its visually distinctive dark lesions featuring characteristic tan centres—a pattern that maps cleanly onto learned convolutional filters. Esca and Gray Mold show marginally lower recall values (0.941 and 0.947, respectively), which reflects genuine inter-class ambiguity: Esca's tiger-stripe chlorosis can superficially resemble early-stage Downy Mildew, and Gray Mold's diffuse grey sporulation sometimes overlaps visually with Powdery Mildew at advanced infection stages. These observations are consistent with diagnostic challenges that human plant pathologists also encounter in field conditions [10]. Of particular practical importance, Healthy vines are classified with near-perfect recall (0.982), which means the system is highly unlikely to raise a false disease alarm—a critical property, since spurious alerts could trigger unnecessary and costly pesticide applications.

D. Ablation Study

The two-stage fine-tuning strategy was validated through controlled ablation. Training the full network end-to-end from epoch one—without the frozen warm-up stage—converged to only 94.1% test accuracy, confirming that the warm-up phase is not merely a convenience but a meaningful contributor to final performance. Removing the Squeeze-and-Excitation attention blocks from the backbone lowered accuracy to 94.6% (−1.8 pp), and omitting CutOut augmentation reduced it to 95.1% (−1.3 pp). Each design choice therefore contributes independently and additively to the overall result.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VI. CONCLUSION AND FUTURE WORK

This paper presented an AI-powered grape disease classification system built on a fine-tuned EfficientNet-B3 backbone trained on a Kaggle grape disease dataset. The proposed framework achieves a top-1 accuracy of 96.4% and a macro F1-score of 0.963 across seven disease categories, outperforming VGG-16, ResNet-50, and MobileNetV2 by margins of 2.7 to 5.2 percentage points. With GPU inference latency of 38 ms and mobile latency of 142 ms under INT8 quantisation, the system satisfies real-time field deployment requirements. The two-stage fine-tuning strategy and domain-adapted augmentation pipeline both contribute meaningfully to the final result, as validated by controlled ablation.

Looking ahead, several avenues are worth pursuing. First, integrating object detection—such as a YOLO-based head—would allow the system to localise disease regions within full-scene vineyard images rather than relying on pre-cropped patches. Second, expanding the dataset to include additional cultivars and geographic viticultural regions would improve domain coverage and model robustness. Third, applying gradient-weighted class activation mapping (Grad-CAM) [17] to generate visual explanations would help agronomists understand which image regions drive each prediction, building trust and facilitating error analysis. Finally, packaging the entire pipeline as a lightweight mobile application with connectivity to agronomic advisory services represents a clear and compelling near-term deployment goal for smallholder viticulturalists.

VII. ACKNOWLEDGMENT

The authors sincerely thank Prof. Soniya Waghmare, Department of Computer Engineering, JSPM's Padmabhooshan Vasantdada Patil Institute of Technology, Pune, for her invaluable guidance, constructive feedback, and continued encouragement throughout this research. The authors also gratefully acknowledge the Kaggle community for curating and maintaining open-access agricultural image datasets that make reproducible research in applied plant pathology possible.

REFERENCES

[1] FAO, "The State of Food and Agriculture 2022," Food and Agriculture Organization of the United Nations, Rome, 2022.
[2] P. Oerke, "Crop losses to pests," Journal of Agricultural Science, vol. 144, no. 1, pp. 31–43, Feb. 2006.
[3] M. Jansen et al., "Connecting disease, drought, and farming losses in Mediterranean viticulture," Plant, Cell & Environment, vol. 41, no. 4, pp. 782–795, 2018.
[4] S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," Frontiers in Plant Science, vol. 7, p. 1419, Sep. 2016.
[5] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv:1704.04861, 2017.
[6] K. Wang et al., "An attention-based deep learning model for plant leaf disease detection," IEEE Access, vol. 9, pp. 120798–120808, 2021.
[7] Y. Chen, H. Yang, and M. Liu, "Few-shot learning for plant disease identification via prototypical networks," Computers and Electronics in Agriculture, vol. 195, p. 106792, Apr. 2022.
[8] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," Computers and Electronics in Agriculture, vol. 145, pp. 311–318, Feb. 2018.
[9] B. Liu et al., "Identification of grape leaf diseases based on deep convolutional neural networks," Symmetry, vol. 10, no. 1, p. 11, Jan. 2018.
[10] P. Jiang et al., "Real-time detection of grape diseases using deep learning based on improved convolutional neural networks," IEEE Access, vol. 7, pp. 59069–59080, 2019.
[11] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Grape crop disease classification using pre-trained deep learning algorithm," Procedia Computer Science, vol. 133, pp. 1040–1047, 2018.
[12] S. Lu, Z. Chen, and C. Tan, "Transfer learning with fine-tuned deep CNNs for plant disease recognition," Biosystems Engineering, vol. 214, pp. 37–52, Feb. 2022.
[13] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, Long Beach, CA, 2019, pp. 6105–6114.
[14] X. Wang et al., "ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks," in Proc. IEEE CVPR, 2017, pp. 2097–2106.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[15] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," Proc. IEEE, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[16] L. Bossard, M. Guillaumin, and L. Van Gool, "Food-101: Mining discriminative components with random forests," in Proc. ECCV, 2014, pp. 446–461.
[17] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE ICCV, 2017, pp. 618–626.
[18] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.
[19] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, May 2015.
[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. NeurIPS, 2012, pp. 1097–1105.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details