



Article

Lightweight and Accurate Deep Learning for Strawberry Leaf Disease Recognition: An Interpretable Approach

Raquel Ochoa-Ornelas ^{1,*}, Alberto Gudiño-Ochoa ², Ansel Y. Rodríguez González ³, Leonardo Trujillo ^{4,5}, Daniel Fajardo-Delgado ¹ and Karla Liliana Puga-Nathal ⁶

¹ Systems and Computation Department, Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Guzmán, Ciudad Guzmán 49100, Jalisco, Mexico; daniel.fd@cdguzman.tecnm.mx

² Electronics and Computing Division, Centro Universitario de Ciencias Exactas e Ingenierías (CUCEI), Universidad de Guadalajara, Guadalajara 44430, Jalisco, Mexico; gudino.ochoa.alberto@gmail.com

³ Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad Académica Tepic, Tepic 63155, Nayarit, Mexico; ansel@cicese.edu.mx

⁴ Electric and Electronics Department, Tecnológico Nacional de México/Instituto Tecnológico de Tijuana, Tijuana 22430, Baja California, Mexico; leonardo.trujillo@tectijuana.edu.mx

⁵ LASIGE, Department of Informatics, Faculty of Sciences, University of Lisbon, 1749-016 Lisbon, Portugal

⁶ Basic Sciences Department, Tecnológico Nacional de México/Instituto Tecnológico de Ciudad Guzmán, Ciudad Guzmán 49100, Jalisco, Mexico; karla.pn@cdguzman.tecnm.mx

* Correspondence: raquel.oo@cdguzman.tecnm.mx

Abstract

Background/Objectives: Strawberry crops are vulnerable to fungal diseases that severely affect yield and quality. Deep learning has shown strong potential for plant disease recognition; however, most architectures rely on tens of millions of parameters, limiting their use in low-resource agricultural settings. This study presents Light-MobileBerryNet, a lightweight and interpretable model designed to achieve accurate strawberry disease classification while remaining computationally efficient for potential use on mobile and edge devices.

Methods: The model, inspired by MobileNetV3-Small, integrates inverted residual blocks, depthwise separable convolutions, squeeze-and-excitation modules, and Swish activation to enhance efficiency. A publicly available dataset was processed using CLAHE and data augmentation, and split into training, validation, and test subsets under consistent conditions. Performance was benchmarked against state-of-the-art CNNs. **Results:** Light-MobileBerryNet achieved 96.6% accuracy, precision, recall, and F1-score, with a Matthews correlation coefficient of 0.96, while requiring fewer than one million parameters (~2 MB). Grad-CAM confirmed that predictions focused on biologically relevant lesion regions. **Conclusions:** Light-MobileBerryNet approaches state-of-the-art performance with a fraction of the computational cost, providing a practical and interpretable solution for precision agriculture.

Keywords: strawberry leaf disease recognition; lightweight CNN; MobileNet



Academic Editors: Ray E. Sheriff and Chiew Foong Kwong

Received: 31 August 2025

Revised: 14 October 2025

Accepted: 15 October 2025

Published: 21 October 2025

Citation: Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; Rodríguez González, A.Y.; Trujillo, L.; Fajardo-Delgado, D.; Puga-Nathal, K.L. Lightweight and Accurate Deep Learning for Strawberry Leaf Disease Recognition: An Interpretable Approach. *AgriEngineering* **2025**, *7*, 355. <https://doi.org/10.3390/agriengineering7100355>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Strawberries are vulnerable to a variety of fungal diseases, including blossom blight (*Botrytis* spp.), gray mold (*Botrytis cinerea*), and powdery mildew (*Podosphaera aphanis*), which can significantly affect both yield and fruit quality. If not promptly identified and controlled, these diseases have the potential to disrupt the strawberry production cycle, leading to inconsistent yields [1–3]. Strawberries rank among the most widely consumed fruits and are extensively cultivated across the globe. Between 2015 and 2021, global annual

production increased by approximately 11.6%, reaching 9.17 million tons in 2021. Early and accurate detection of these diseases is therefore essential for effective management and timely intervention [1].

Deep learning, particularly convolutional neural networks (CNNs), has demonstrated great potential in plant disease recognition due to their ability to automatically learn and extract relevant features from images [4–6]. Transfer learning, which consists of adapting a pre-trained model to a new dataset, is especially valuable in agricultural applications, where labeled data are often scarce [4–8]. This approach has been successfully applied to detect various plant diseases, including those affecting strawberries [1,2]. The integration of CNNs with transfer learning techniques represents a promising strategy to optimize disease management and improve agricultural productivity. Furthermore, the development of lightweight and interpretable models enables high accuracy and efficiency, facilitating deployment in real-world precision agriculture scenarios.

The adoption of such models offers multiple benefits, including significant economic impact, since strawberry diseases can cause substantial financial losses by reducing yield and quality [1–3]. Deep learning-based disease detection also allows optimization of cultivation practices, enabling early interventions and reducing pesticide use [2,9]. Moreover, transfer learning-based models require fewer data and less computational power for training, making them well suited for real-time applications in resource-limited environments [4–6,8,10,11].

Several deep learning models, particularly those based on transfer learning, such as VGG16, EfficientNetV2B1, and MobileNetV2, have been used for the identification and classification of various strawberry diseases, achieving high accuracy [1,2,12]. Transfer learning has shown remarkable potential to improve performance in disease detection even with small datasets, making it a reliable and adaptable solution for automating strawberry disease diagnosis [1,2,12]. In this context, MobileNetV2 stands out for its ability to achieve a balance between accuracy and efficiency, making it particularly suitable for agricultural environments where computational resources are limited [2].

Transfer learning models must be adaptable to different types of diseases, including those with visually similar symptoms. Fine-tuning pre-trained models on datasets containing specific diseases has been shown to significantly improve performance. However, achieving high accuracy in scenarios involving a variety of diseases remains challenging [13,14]. One of the main difficulties lies in the variability between the source and target domains, influenced by factors such as lighting, background, and strawberry cultivar [15]. Approaches using hyperspectral images and spectral signature analysis have been explored to enhance early detection, but these methods require specialized equipment and technical expertise [16]. However, most existing studies optimize accuracy over limited disease categories and overlook the trade-off between model compactness and generalization.

Although deep learning models have achieved strong performance, most of them rely on large, high-parameter architectures—such as EfficientNet, Inception, or ResNet—that demand substantial computational resources and are difficult to deploy in real agricultural environments [10–14]. Conversely, lightweight models often sacrifice accuracy or robustness, particularly when generalizing across multiple visually similar disease categories. Thus, an evident research gap persists in developing lightweight yet generalizable models that retain high diagnostic accuracy while remaining feasible for real-world agricultural deployment [10].

To address these challenges, we propose a lightweight and interpretable architecture specifically designed for strawberry disease detection. This work directly addresses that gap by proposing a scalable, interpretable CNN tailored for practical deployment in strawberry disease monitoring systems. The proposed model directly targets the above limitations by

pursuing a trade-off between performance and efficiency, offering a scalable solution for mobile and edge computing scenarios. It incorporates advanced techniques inspired by architectures such as MobileNetV2, which has demonstrated high accuracy with a reduced number of parameters [2,10]. The approach also leverages pre-trained models on large datasets, subsequently fine-tuned on strawberry disease images to accelerate training and improve accuracy [1,2]. The goal is to achieve classification accuracy comparable to or surpassing current models [2,10], while ensuring computational efficiency for deployment on mobile and edge devices [10,11].

The main contributions of this work are as follows:

- Development of a novel lightweight deep learning architecture, inspired by MobileNetV3Small, integrating Inverted Residual blocks with depthwise separable convolutions, Squeeze-and-Excitation modules, and explicit Swish activation to enhance feature extraction efficiency while maintaining low computational cost.
- Significant reduction in model complexity, achieving a smaller number of parameters and reduced model size compared to baseline architectures such as MobileNetV3Small, MobileNetV3Large, and EfficientNetB0, enabling deployment on low-cost mobile and edge devices.
- Interpretability through visual attention mechanisms (e.g., Grad-CAM) to highlight the region's most influential in the model's decision-making process.

The remainder of this paper is organized as follows: Section 2 reviews related works on plant disease detection, lightweight CNN architectures, and mobile deployment strategies. Section 3 describes the dataset, preprocessing pipeline, and the proposed lightweight CNN model. Section 4 presents the experimental results and comparative analysis with other transfer learning architectures. Section 5 discusses the implications of the findings, potential limitations, and opportunities for improvement. Finally, Section 6 concludes the paper and outlines directions for future research.

2. Related Work

Deep learning architectures, particularly CNNs, have become a dominant approach for strawberry disease detection, achieving notable success in classifying diseases such as gray mold, powdery mildew, and blossom blight [1,10,12,17–20]. Sijan Karki et al. [1] reported a 94.4% accuracy using ResNet-50 with feature extraction, specifically for Powdery Mildew Leaf, Gray Mold Leaf, and healthy leaf classification. Improvements to AlexNet through fine-tuning, aimed at reducing the number of trainable parameters and shortening training time, have yielded accuracies of up to 97.35% [1].

BerryNet-Lite, proposed by Jianping Wang et al. (2024), integrates an efficient channel attention (ECA) mechanism with a multilayer perceptron (MLP), achieving an accuracy of 99.45%—surpassing ResNet34, VGG16, and AlexNet [10]. Transfer learning has proven particularly advantageous, significantly reducing training time and trainable parameters while maintaining or even improving accuracy, a benefit that becomes more pronounced in scenarios with limited datasets [1,19].

Among pre-trained models, MobileNetV2 has achieved accuracies of 98.97% in recognizing gray mold, powdery mildew, and blossom blight [2], demonstrating a consistent balance between accuracy and computational efficiency for agricultural applications. ResNet-50, when fine-tuned, has also produced accuracies above 94.4% for strawberry disease identification [2,17,20]. DenseNet-121, after fine-tuning, achieved 94.1% accuracy, confirming its suitability for detecting diverse strawberry diseases [1,17]. EfficientNetB0–B7 variants have achieved benchmark performances exceeding 99% for strawberry disease classification, further reinforcing the potential of transfer learning for high-accuracy detection in this domain [12]. Given their balance of efficiency and accuracy, MobileNetV2

and EfficientNetB0 are considered particularly suitable for real-time monitoring and mobile deployment [2,12]. Although most studies address gray mold and powdery mildew explicitly, blossom blight is often included implicitly within broader disease classification tasks using MobileNetV2 and ResNet-50, indicating that these architectures can also handle its recognition [1,2]. Recently, several transformer-based architectures have been introduced for strawberry disease recognition. Li et al. [21] proposed a Spatial Convolutional Self-Attention Transformer (SCSA-Transformer) combining convolutional encoding and multi-head self-attention to enhance disease localization in complex backgrounds, achieving 99.1% accuracy. An optimized strawberry disease and quality detection using Vision Transformers and attention-based CNN hybrids, reporting accuracies above 98% [22]. Likewise, Nguyen et al. [23] demonstrated the potential of transfer-learned Vision Transformers for seven-class strawberry disease identification, achieving F1-scores near 0.93.

Beyond conventional classification models, alternative approaches have emerged. Mingzhou Chen et al. (2025) explored YOLOv8-based segmentation for detecting leaf lesions and powdery mildew, outperforming SOLOv2, YOLACT, and YOLOv7-seg with 92% segmentation accuracy, 85.2% recall, and 90.4% mean average precision [24]. Yousef Alhwaiti et al. (2025) compared YOLOv3 and YOLOv4, reporting 97% accuracy and 92% mAP for YOLOv3, while YOLOv4 demonstrated lower complexity, faster inference, and improved precision, achieving 98% accuracy and 98% mAP [25].

Ensemble learning has also been investigated. Haram Kim et al. (2023) combined RegNet and EfficientNet for leaf disease classification, obtaining an 85% accuracy—below expectations for strawberry leaf diseases [26]. Hyperspectral imaging, when combined with CNNs, can integrate multi-dimensional features such as spectral fingerprints and vegetation indices, achieving recognition accuracies between 88.9% and 96.6% [27]. Nevertheless, these methods require specialized equipment and substantial computational resources, limiting their practicality for on-field deployment.

Despite these advances, deep learning and transfer learning methods for strawberry disease recognition still face critical challenges. High-accuracy recognition often demands large-scale datasets [28], while variability in image acquisition—due to differences in lighting, background, and cultivar—complicates feature extraction, leading to degraded performance under complex conditions [29,30]. Even with transfer learning and advanced deep neural networks, classification performance can be hindered by domain shifts and dataset imbalance [31].

In this context, our proposed lightweight CNN addresses several of these limitations. It is inspired by MobileNetV3Small but incorporates design refinements—including Inverted Residual blocks with depthwise separable convolutions, integrated Squeeze-and-Excitation modules, and explicit Swish activation—to enhance feature representation without sacrificing efficiency. The architecture is designed to achieve a balance between accuracy and computational efficiency, making it suitable for deployment on low-cost mobile and edge devices. By combining architectural efficiency with interpretability (via Grad-CAM visualizations), the model not only delivers competitive accuracy but also provides transparent decision-making, addressing both performance and usability gaps identified in the current literature.

3. Materials and Methods

3.1. Image Datasets

The experiments in this study were conducted using the Strawberry Disease Detection Dataset (Instance Segmentation Dataset for Seven Types of Strawberry Diseases), which contains a total of 2500 images accompanied by segmentation annotation files for seven distinct strawberry disease categories. The dataset was compiled by the Artificial Intelligence Laboratory at the Department of Computer Science and Engineering, Jeonbuk National

University (JBNU), and aggregates imagery from multiple reputable institutions and public repositories, including the University of Bologna, the USDA Agricultural Research Service (ARS), the University of California Agriculture and Natural Resources, Cornell University, the University of Kentucky, and other academic and governmental sources [32].

According to Afzaal et al. [32], approximately 80% of the images were collected in real greenhouse environments in South Korea using camera-equipped mobile phones under natural illumination, while the remaining 20% were obtained from open-access online agricultural repositories to increase variability in background, lighting, and disease stages. The dataset therefore includes both close-up and distant perspectives and covers different phases of disease progression, contributing to a realistic visual diversity that supports model robustness and generalization to field conditions.

The dataset comprises seven disease categories: angular leaf spot, anthracnose fruit rot, blossom blight, gray mold, leaf spot, powdery mildew (fruit), and powdery mildew (leaf). These diseases represent critical threats to strawberry production, affecting flowers and fruits during key developmental stages and causing substantial economic losses if left untreated. Representative examples of each class are shown in Figure 1.

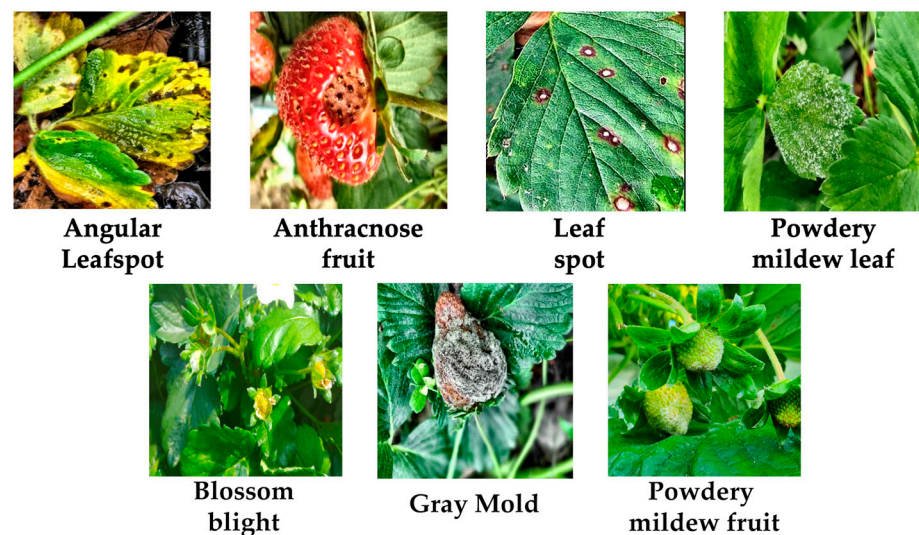


Figure 1. Representative examples of the seven strawberry disease classes used in this study extracted from the strawberry disease detection dataset.

In this work, the segmentation annotations provided in the original dataset were not used, as the main objective was to develop an image-level classification model rather than a segmentation-based approach. This design choice simplifies deployment and reflects realistic field scenarios where pixel-wise labels are typically unavailable. All images were preprocessed and augmented as described in Section 3.2 to enhance robustness under diverse illumination conditions, background variability, and field-level complexity.

3.2. Preprocessing

All images were first resized to a fixed resolution of 224×224 pixels to ensure consistency with the input requirements of the proposed lightweight CNN architecture. Images were converted to RGB format when necessary and stored in category-specific directories corresponding to the selected disease classes.

To improve contrast and enhance visual details relevant for disease recognition, we applied Contrast Limited Adaptive Histogram Equalization (CLAHE) to all images prior to augmentation [31,33]. For RGB images, CLAHE was applied to the luminance channel after conversion to the YUV color space, with a clip limit of 2.0 and a tile grid size of

8×8 . This step helped normalize lighting variations and highlight disease symptoms under varying field conditions.

Following contrast enhancement, data augmentation was performed applying a diverse set of transformations to improve the model's robustness against changes in orientation, scale, and illumination. The augmentation operations included:

- Random rotations of up to $\pm 25^\circ$ to simulate different camera angles;
- Zoom scaling within a range of 0.8 to 1.2 to account for variable shooting distances;
- Horizontal and vertical flips to replicate different leaf and fruit orientations in the field;
- Brightness adjustments within the range [0.5, 1.5] to handle variations in natural lighting;
- Nearest-neighbor filling for pixels introduced during geometric transformations.

Augmentation was applied iteratively until each disease class reached a target of 1000 images, ensuring a balanced dataset across all categories and reducing the risk of overfitting [34–38]. This approach mitigated potential class imbalance, which can adversely affect model training and evaluation. During this process, augmented images were also processed with CLAHE to maintain contrast enhancement consistency between original and synthetic samples.

3.3. Proposed Lightweight CNN Model

To address the computational and deployment constraints of in-field strawberry disease detection, we propose Light-MobileBerryNet, a lightweight convolutional neural network inspired by MobileNetV3-Small but re-designed with the following traits:

1. A key efficiency lever, which involves is the use of depthwise separable convolutions in inverted-residual (IR) blocks;
2. The ability to improve non-linear representation using the Swish activation [36,38];
3. The ability to minimize architectural redundancy by carefully selecting expansion factors, strides, and block depths to preserve accuracy while reducing parameters and multiply–accumulate operations (MACs) [39].

These choices are motivated by the need to detect fine-grained, texture-like cues (e.g., mold hyphae, lesion borders) typical of strawberry leaf disease, under edge latency and memory budgets [1,10]. The overall architecture is summarized in Figure 2.

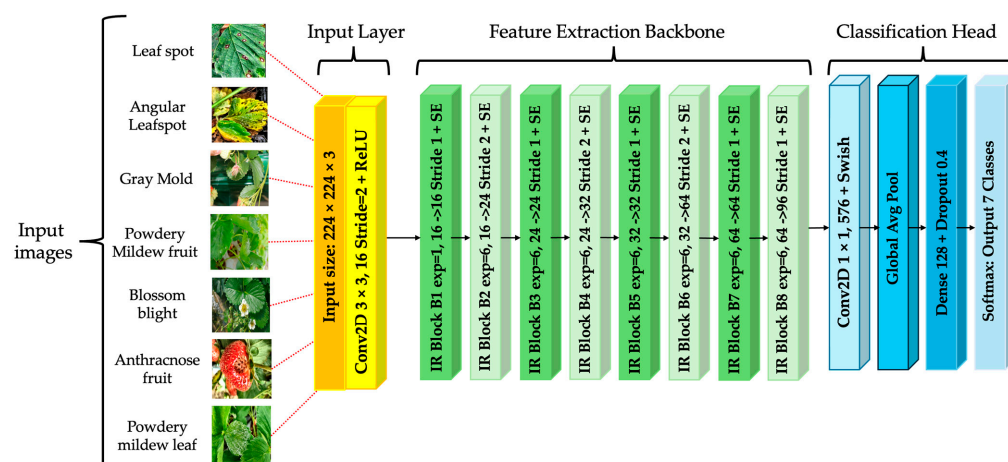


Figure 2. Schematic overview of the proposed Light-MobileBerryNet architecture for seven-class strawberry disease recognition.

A key efficiency lever involves the use of depthwise separable convolutions in inverted-residual (IR) blocks [40,41]. For an input with M channels and N output channels using a $K \times K$ kernel, the parameter/MAC count of a standard convolution is:

$$Params_{std} = K^2 \cdot M \cdot N \quad (1)$$

where a depthwise + pointwise decomposition uses

$$Params_{dw+pw} = K^2 \cdot M + M \cdot N \quad (2)$$

Equation (2) yields a substantial reduction when $K \ll \min(M, N)$. We further retain representational capacity by expanding to tM channels (expansion factor t) before the depthwise operation and projecting back to N channels. The Swish activation function is defined as

$$Swish(x) = x\sigma(\beta x), \quad (\beta = 1 \text{ in this work}) \quad (3)$$

where x denotes the input activation, σ is the logistic sigmoid function, and β is a scaling parameter that controls the smoothness of the non-linearity [39]. Channel recalibration is achieved using SE attention, where global average pooling is applied to each feature map to obtain channel descriptors z_c which are then reweighted through a bottleneck and sigmoid gating [40]:

$$z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W x_{ijc}, \quad s = \sigma(W_2 \delta(W_1 z)), \quad y_{ijc} = s_c x_{ijc}, \quad (4)$$

where x_{ijc} denotes the activation at spatial location (i, j) and channel c , H and W are the spatial height and width of the feature map, z_c is the aggregated channel descriptor obtained by global average pooling, W_1 and W_2 are the learnable weight matrices of the SE bottleneck, δ is the rectified linear unit (ReLU) activation, σ is the logistic sigmoid function, s_c is the channel-wise scaling factor, y_{ijc} is the recalibrated output activation at position (i, j, c) , and r is the reduction ratio of the bottleneck (we use $r = 4$) controlling the SE overhead. To facilitate quantization and pruning, the architecture was kept shallow-to-moderate (eight IR blocks), avoiding exotic operations and concentrating capacity where it benefits most, namely mid- and high-level texture cues.

3.3.1. Architecture of Light-MobileBerryNet

Given an RGB input $X \in \mathbb{R}^{224 \times 224 \times 3}$, the network applies a 3×3 stem convolution (stride 2, 16 filters) with batch normalization and ReLU, followed by eight inverted-residual blocks. A stack of eight inverted-residual blocks was used, and each block contains an optional 1×1 expansion to tM channels, 3×3 depthwise convolution with stride, SE module, and 1×1 projection to N output channels; Swish activations are used after expansion and depthwise. A skip connection is used when $s = 1$ and $M = N$. The stack ends with a 1×1 convolution (576 channels) with Swish, global average pooling, a dense layer (128 units, ReLU, L_2 regularization, dropout 0.4), and a final softmax over the seven classes. Table 1 summarizes the macro-architecture. Here, t is the expansion factor, C_{in} and C_{out} are input/output channels of the block, and SE indicates Squeeze-and-Excitation usage.

Table 1. Architecture of Light-MobileBerryNet.

Stage	Operator	Exp. t	$C_{in} \rightarrow C_{out}$	Kernel	Stride	SE
Stem	Conv-BN-ReLU	-	7→16	3×3	2	-
B1	IR (Swish)	1	16→16	3×3	1	✓
B2	IR (Swish)	6	16→24	3×3	2	✓
B3	IR (Swish)	6	24→24	3×3	1	✓
B4	IR (Swish)	6	24→32	3×3	2	✓
B5	IR (Swish)	6	32→32	3×3	1	✓
B6	IR (Swish)	6	32→64	3×3	2	✓
B7	IR (Swish)	6	64→64	3×3	1	✓
B8	IR (Swish)	6	64→96	3×3	1	✓
Head	Conv1 × 1-Swish	-	96→576	1×1	1	-
	GAP + FC + Drop	-	576→128	-	-	-
	Classifier (Softmax)	-	128→7	-	-	-

Skip if $s = 1$ and $C_{in} = C_{out}$.

3.3.2. Layers of Light-MobileBerryNet

The computational flow of Light-MobileBerryNet starts with a stem convolution that extracts low-level edges and textures. Specifically, the input image $\mathbf{X} \in \mathbb{R}^{224 \times 224 \times 3}$ is processed with a 3×3 convolution (stride 2, 16 filters), followed by batch normalization and the ReLU activation, producing the initial feature representation $\mathbf{H}_0$:

$$\mathbf{U} = \text{BN}(\mathbf{X} \cdot \mathbf{W}_{3 \times 3}^{(16)}), \quad \mathbf{H}_0 = \text{ReLU}(\mathbf{U}) \quad (5)$$

The backbone consists of eight IR blocks. Each block follows the general principle of expansion, depthwise convolution, channel recalibration via SE (Equation (4)), and projection. Formally, given an input $\mathbf{H} \in \mathbb{R}^{H \times W \times M}$, the expansion step increases the channels to tM with a 1×1 convolution and Swish activation (Equation (3)) [41]:

$$\mathbf{E} = \varnothing \left(\text{BN} \left(\mathbf{H} \cdot \mathbf{W}_{1 \times 1}^{tM} \right) \right) \quad (6)$$

This is followed by a 3×3 depthwise convolution applied channel-wise, also activated with Swish:

$$\mathbf{D} = \varnothing \left(\text{BN} \left(\mathbf{E} \odot \mathbf{K}_{3 \times 3} \right) \right) \quad (7)$$

where $\odot$ denotes depthwise convolution. The output $\mathbf{D}$ is then recalibrated through the SE mechanism already described in Equation (4), yielding channel-wise scaling coefficients s_c that modulate the activations [39]. The block concludes with a 1×1 projection to N output channels:

$$\mathbf{P} = \text{BN} \left(\mathbf{Y} \cdot \mathbf{W}_{1 \times 1}^{(N)} \right) \quad (8)$$

When the stride is one and the channel dimensions match ($s = 1, M = N$), a residual skip connection is applied, producing

$$\mathbf{H}' = \begin{cases} \mathbf{P} + \mathbf{H}, & \text{if } s = 1 \wedge M = N, \\ \mathbf{P}, & \text{otherwise} \end{cases} \quad (9)$$

After the IR stack, the head of the network expands the feature representation to 576 channels with a 1×1 convolution and Swish activation, followed by global average pooling to obtain a compact vector $\mathbf{g}$:

$$\mathbf{Z} = \varnothing \left(\text{BN} \left(\mathbf{H}' \cdot \mathbf{W}_{1 \times 1}^{(576)} \right) \right), \quad \mathbf{g} = \text{GAP}(\mathbf{Z}) \quad (10)$$

The classification stage projects g into a 128-dimensional dense representation with ReLU activation, applies dropout regularization, and finally maps to the seven output classes with a softmax layer:

$$h = \delta(W_{fc}g + b_{fc}), \tilde{h} = \text{Dropout}_{0.4}(h), \hat{y} = \text{Softmax}(W_c\tilde{h} + b_c) \quad (11)$$

The model is optimized by minimizing the categorical cross-entropy loss with an additional $\uparrow_2$ regularization term applied to the fully connected layer:

$$\mathcal{L} = -\sum_{k=1}^3 y_k \log \hat{y}_k + \lambda \|W_{fc}\|_2^2 \quad (12)$$

where $y \in \{0, 1\}^3$ is the one-hot encoded ground truth vector, $\hat{y}$ is the predicted probability distribution, and $\lambda = 0.01$.

3.4. Visual Saliency Maps with Grad-CAM

To enhance interpretability and provide insights into the decision-making process of Light-MobileBerryNet, we employed Gradient-weighted Class Activation Mapping (Grad-CAM). This technique allows us to visualize the regions of strawberry leaves and fruits that most strongly contribute to the classification decision, thereby ensuring that the network is focusing on disease-specific lesions such as mold hyphae, blight-affected tissues, or powdery patches, rather than irrelevant background cues. The Grad-CAM approach has proven particularly useful in lightweight CNNs, where interpretability is crucial for validating model robustness in real agricultural applications [30,42,43].

Formally, let y_c denote the pre-softmax score associated with class c , and let A^k represent the k -th feature map of the last convolutional layer, with spatial indices (i, j) . Grad-CAM first computes the gradients of y_c with respect to A_{ij}^k , propagating the class-specific signal back to the convolutional layer. These gradients are then global average pooled across the spatial dimensions to produce the neuron importance weights α_c^k :

$$\alpha_c^k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y_c}{\partial A_{ij}^k} \quad (13)$$

where $Z = H \times W$ is the product of the height and width of the feature map. The weights α_c^k quantify the relative contribution of each feature map A^k to the prediction of class c .

The Grad-CAM saliency map $L_c^{\text{Grad-CAM}} \in R^{u \times v}$ is then obtained as the weighted sum of the feature maps followed by a ReLU:

$$L_c^{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_c^k A^k\right) \quad (14)$$

here, ReLU is defined as

$$\text{ReLU}(x) = \begin{cases} x, & x > 0, \\ 0, & x \leq 0, \end{cases} \quad (15)$$

Ensuring that only the positive contributions supporting the prediction of class c are preserved, Grad-CAM highlights the discriminative regions in the input image that drive the final classification decision. When applied to the strawberry disease dataset, Grad-CAM demonstrates that Light-MobileBerryNet attends to biologically relevant regions. This visual interpretability provides an additional layer of validation, reinforcing the reliability of the proposed lightweight model for real-world deployment in precision agriculture [43].

4. Results

All models, including the proposed Light-MobileBerryNet and the baseline architectures, were implemented in Python 3.13 and trained under identical hyperparameter settings to ensure a fair comparison. Experiments were executed on a workstation equipped with an Intel i9-13900KF CPU (3.0 GHz, 64 GB RAM), an NVIDIA RTX A4500 GPU (20 GB VRAM), and a 1 TB PCIe SSD.

The reported latency, floating-point operations (FLOPs), and multiply–accumulate operations (MACs) represent per-image inference on this desktop hardware. These metrics serve primarily for reproducibility and benchmarking, not as direct indicators of mobile-device performance. Nevertheless, given the small computational footprint (~2 MB model size), the architecture is expected to achieve real-time inference on mid-range mobile SoCs or embedded GPUs (e.g., NVIDIA Jetson Nano [44]).

The dataset was divided into sections (80% training, 10% validation, and 10% testing) using a fixed random seed to preserve reproducibility (~700 test images). Cross-validation was initially considered, but a fixed validation subset was preferred to reduce computational cost—an approach consistent with common computer-vision practices [45]. Images were resized to 224×224 RGB and fed through Keras data generators (batch = 16, shuffling enabled for training/validation only).

Training used the Adam optimizer (learning rate = 1×10^{-3}) and categorical cross-entropy loss, capped at 30 epochs. Performance typically plateaued beyond this point, as observed consistently across the proposed and baseline architectures. Extending the training beyond 30 epochs did not yield further improvement and, in some models, even led to minor degradation due to overfitting. Early stopping (patience = 5), learning-rate reduction on plateau (factor = 0.5, patience = 3), and model checkpointing were employed to prevent overfitting and ensure generalization.

The evaluation encompassed Light-MobileBerryNet—designed as a lightweight MobileNetV3-Small-inspired architecture—and a comprehensive suite of state-of-the-art CNNs, including MobileNetV2/V3-Large, EfficientNet (B0/B3/B7, V2-S/M/L), DenseNet (121/169), NASNetMobile, ResNet (50/101), InceptionResNetV2, and Xception. All models shared identical optimization and early stopping protocols, ensuring that performance differences reflected architectural design rather than training conditions.

Finally, an ablation study was performed to quantify the contribution of Light-MobileBerryNet key components, including the Swish activation, the number of inverted residual blocks, and the expansion factor.

4.1. Performance Metrics

To assess model performance comprehensively, we adopted a set of evaluation metrics commonly used in plant disease recognition; These include accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and the area under the receiver operating characteristic curve (ROC-AUC). Accuracy measures the overall proportion of correctly classified samples, while precision quantifies the proportion of predicted positives that are true positives, capturing the model's ability to avoid false alarms. Recall assesses sensitivity, i.e., the proportion of true positives correctly identified, and the F1-score harmonically balances precision and recall, particularly relevant under imbalanced distributions. MCC provides a more stringent correlation-based evaluation that incorporates all elements of the confusion matrix and is particularly informative in multiclass settings. ROC-AUC complements these by evaluating class separability independent of specific thresholds,

highlighting the discriminative ability of the model [1,32,46,47]. The metrics were computed according to the following definitions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (17)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (18)$$

$$F_1\text{score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (19)$$

$$\text{ROC AUC} = \int_0^1 \text{TPR}(FPR) d(FPR) \quad (20)$$

$$\text{MCC} = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (21)$$

where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives, respectively. TPR are True Positive Rate and FPR are False Positive Rate. Together, these indicators provide both a global and class-level understanding of performance, balancing overall accuracy with the model's capacity to identify each disease category reliably and consistently.

4.2. Model Evaluation

The learning dynamics of Light-MobileBerryNet (Figure 3) reveal a rapid and stable convergence. Training and validation losses declined sharply during the initial epochs and stabilized around epoch 22. In parallel, accuracy increased rapidly, with validation performance plateauing near 97% after the tenth epoch, while training accuracy continued to improve, reaching its peak around epoch 29 in both loss and accuracy curves. This behavior indicates an efficient optimization process, where regularization prevented overfitting and ensured that performance gains were not achieved at the expense of generalization.

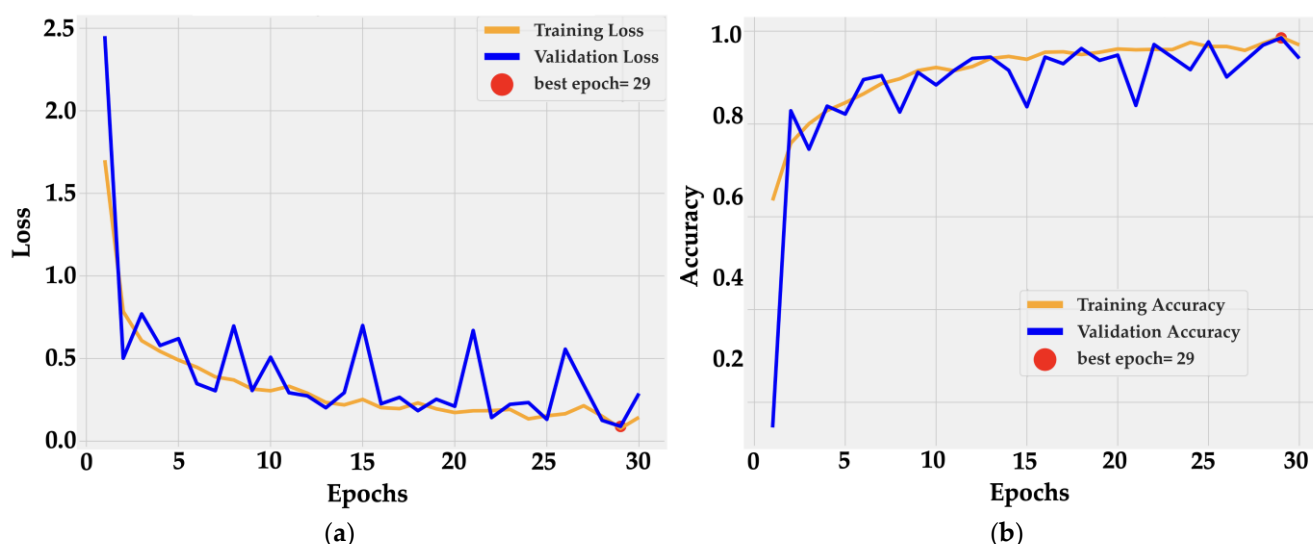


Figure 3. Training dynamics of Light-MobileBerryNet over 30 epochs. (a) Training and validation loss curves; (b) Training and validation accuracy curves.

The confusion matrices in Figure 4 show consistent performance across the seven disease categories, with only 24 misclassifications out of 700 test images (around 3.4%).

Most errors arose from confusions between powdery mildew fruit and gray mold, along with isolated cases involving angular leafspot, anthracnose fruit rot, leaf spot, and powdery mildew leaf.

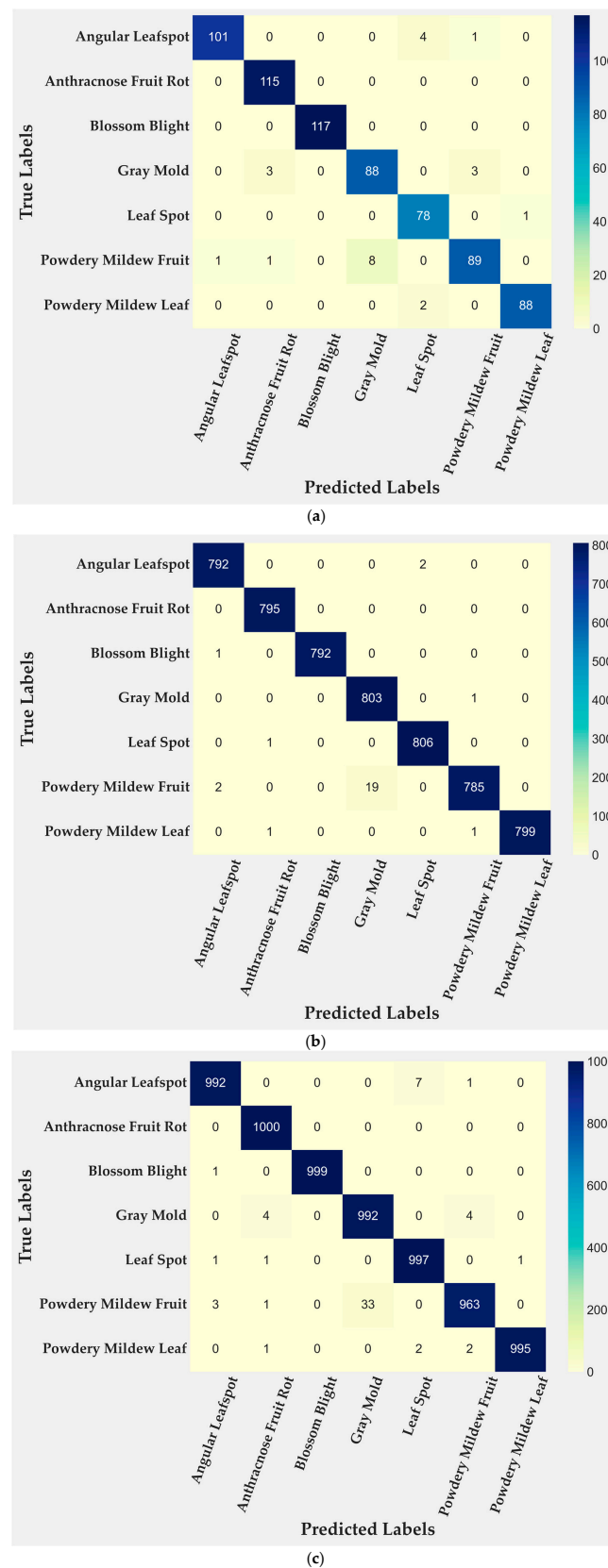


Figure 4. Confusion matrices for Light-MobileBerryNet across different evaluation settings. (a) Test set; (b) train set; (c) all datasets.

On the training set, the error pattern mirrored that of the test data, with misclassifications concentrated between powdery mildew fruit and gray mold. In contrast, Anthracnose fruit rot was classified perfectly. Out of 5600 training images, only 28 were misclassified, corresponding to an error rate of 0.5%.

When evaluated on the entire dataset, the model reproduced the same error pattern, with powdery mildew fruit accounting for most of the 33 misclassifications and gray mold showing the highest number of false positives and negatives. Other categories presented between 1 and 8 errors, except for anthracnose fruit rot, which was classified perfectly, and blossom blight, with only a single error. In total, 62 misclassifications were observed out of 7000 images, corresponding to an error rate of around 0.9%.

Table 2 summarizes the class-wise performance in terms of precision, recall, and F1-score across the test, training, and full datasets. We emphasize recall and F1 because together they capture the model's ability to minimize both false negatives and false positives, offering a more reliable assessment than accuracy alone. On the test set, the model achieved perfect recognition for one class (F1 of 1.000), while the most challenging were gray mold (F1 of 0.926) and powdery mildew fruit (F1 of 0.927), reflecting higher false positives in the former and higher false negatives in the latter. All other categories maintained an F1-score above 0.95. On training data, performance was nearly flawless (F1 between 0.985 and 0.999; recall between 0.973 and 1.000), and on the full dataset it remained stable (F1 no lower than 0.977; recall between 0.963 and 1.000). These results confirm the balanced control of both error types across categories, underscoring the robustness of the proposed model.

Table 2. Class-wise performance (Precision, Recall, and F1-score) for the test, training, and full dataset.

Dataset	Class	Precision	Recall	F1-Score
Test	Blossom Blight	1.000	1.000	1.000
	Gray Mold	0.917	0.936	0.926
	Powdery Mildew Fruit	0.957	0.898	0.927
	Angular Leafspot	0.990	0.952	0.971
	Anthracnose Fruit Rot	0.967	1.000	0.983
	Powdery Mildew Leaf	0.989	0.978	0.983
	Leaf Spot	0.929	0.987	0.957
Train	Blossom Blight	1.000	0.998	0.999
	Gray Mold	0.977	0.998	0.987
	Powdery Mildew Fruit	0.997	0.973	0.985
	Angular Leafspot	0.996	0.997	0.996
	Anthracnose Fruit Rot	0.997	1.000	0.998
	Powdery Mildew Leaf	1.000	0.997	0.998
	Leaf Spot	0.997	0.998	0.998
All	Blossom Blight	1.000	0.999	0.999
	Gray Mold	0.967	0.992	0.979
	Powdery Mildew Fruit	0.992	0.963	0.977
	Angular Leafspot	0.994	0.992	0.993
	Anthracnose Fruit Rot	0.993	1.000	0.996
	Powdery Mildew Leaf	0.998	0.995	0.996
	Leaf Spot	0.991	0.997	0.994

The global indicators in Table 3 confirm the consistency of the model across datasets. On the test set, accuracy, precision, recall, and F1-score were all 96.6%, with an MCC of 0.960, confirming balanced predictive performance across disease categories. On the training set, performance approached perfection (accuracy 99.5%, F1-score 0.994, MCC 0.994), and evaluation on the full dataset maintained this stability (accuracy 99.1%, F1-score

0.991, MCC 0.990). These results highlight not only the strong discriminative capacity of the proposed architecture but also its reliability under realistic, data-intensive conditions.

Table 3. Global evaluation metrics for the test, training, and full dataset.

Dataset	Accuracy	Precision	Recall	F1-Score	MCC
Test	0.966	0.966	0.966	0.966	0.960
Train	0.995	0.995	0.995	0.994	0.994
All	0.991	0.991	0.991	0.991	0.989

The receiver operating characteristic (ROC) and precision–recall (PR) curves in Figure 5 further confirm the discriminative strength of the proposed model for strawberry leaf disease recognition. All classes achieved near-perfect separability, with area under the curve (AUC) values above 0.996 in the ROC analysis. blossom blight reached an AUC of 1.000, reflecting its highly distinctive symptom patterns, while gray mold (AUC = 0.9962) was the lowest, consistent with its visual similarity to powdery mildew fruit. In the PR curves, the trend was comparable, with most classes exceeding an AUC of 0.998. The only exceptions were gray mold (AUC = 0.9824) and powdery mildew fruit (AUC = 0.9790), which exhibited minor drops in precision at high recall levels, again underscoring their overlap in visual features. Overall, the uniformly high AUC values across both metrics demonstrate not only excellent sensitivity and specificity but also a strong balance between precision and recall, reinforcing the robustness of Light-MobileBerryNet for strawberry leaf disease recognition.

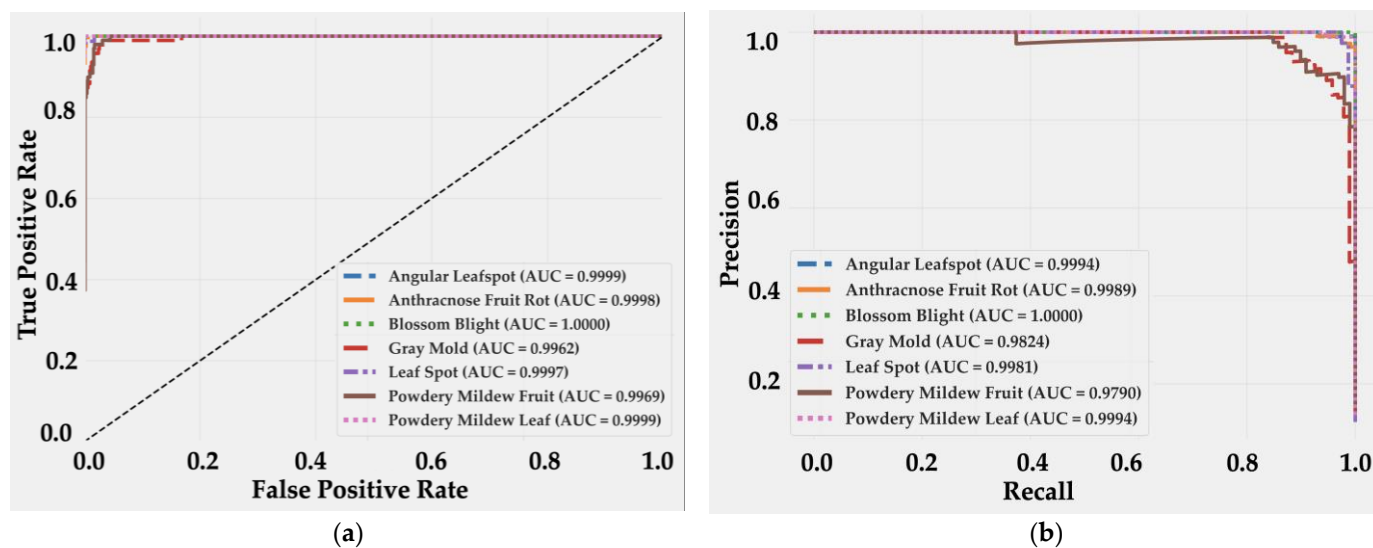


Figure 5. ROC and precision–recall curves for Light-MobileBerryNet on the test set. (a) Receiver operating characteristic curves; (b) Precision-Recall curves.

4.3. Ablation Study

To quantify the influence of key architectural choices, an ablation study was performed in which each component of the proposed configuration—Swish activation, eight inverted residual (IR) blocks, and an expansion factor of 6—was independently modified. Table 4 summarizes the computational efficiency and predictive performance of each variant. The metrics include the number of parameters, model size, MACs, FLOPs, throughput (frames per second, FPS), and latency, as well as standard classification metrics. All MACs, FLOPs, and latency values refer to per-image measurements.

Table 4. Ablation study of the proposed architecture. The MACs and FLOPs results are per image.

Architecture	Params (M)	Size (MB)	MACs (M)	FLOPs (M)	FPS	Latency (ms)	Accuracy	Precision	Recall	F1-Score	MCC
Proposed (Swish, 8 IR blocks)	0.532	2.03	143.97	287.94	131.85	7.58	0.966	0.966	0.966	0.966	0.960
ReLU instead of Swish	0.533	2.03	139.89	279.78	157.43	6.35	0.901	0.913	0.901	0.903	0.886
7 IR blocks instead of 8	0.372	1.42	127.16	254.33	124.75	8.02	0.957	0.961	0.957	0.958	0.950
Expansion factor 4 instead of 6	0.353	1.35	91.21	182.43	159.26	6.28	0.941	0.944	0.941	0.942	0.934

Replacing Swish with ReLU maintained a comparable theoretical workload (≈ 280 M FLOPs per image) and slightly improved throughput (157 FPS vs. 132 FPS) yet produced a sharp performance decline—accuracy decreased by 6.5% and MCC by 0.074. This confirms that the smoother nonlinearity of Swish is essential for capturing subtle lesion textures and non-linear feature interactions. Reducing the network to seven IR blocks decreased parameters by $\approx 30\%$ and model size from 2.03 MB to 1.42 MB, while slightly lowering accuracy (-0.9%) and MCC (-0.01). The additional IR block therefore contributes non-redundant representational depth with minimal computational overhead.

Lowering the expansion factor from 6 to 4 yielded the smallest and fastest variant (1.35 MB, 159 FPS, 6.3 ms per-image latency) but also the lowest reliability (accuracy -2.2% , MCC -0.026). The narrower bottlenecks limit channel capacity and feature abstraction despite their speed advantage. Overall, the proposed configuration offers the most favorable balance: it is compact (~ 2 MB), efficient (~ 132 FPS, 7.6 ms per image), and achieves the highest predictive consistency (accuracy = 96.6%, MCC = 0.960). These results demonstrate that Swish activation, eight IR blocks, and an expansion factor of 6 act synergistically, providing an optimal trade-off between accuracy and efficiency for lightweight strawberry disease recognition.

4.4. Comparative Analysis with State-of-the-Art Models

Table 5 summarizes the comparative performance of Light-MobileBerryNet against a broad suite of state-of-the-art convolutional architectures, all trained under identical hyperparameter, optimization, and early stopping settings.

Across the benchmark, Light-MobileBerryNet attains 96.6% test accuracy and an MCC of 0.960 while requiring only 0.53 million parameters (~ 2 MB). This translates into a $>99\%$ reduction in size compared with EfficientNet-V2L (118 M parameters, 450 MB) and only a 0.01 drop in accuracy. In practical terms, the proposed network retains almost the full discriminative capacity of the most complex EfficientNet-V2 models while remaining lightweight enough for real-time inference on mobile hardware (~ 7.6 ms per image).

When compared with MobileNetV3-Small—its most conceptually similar baseline—Light-MobileBerryNet achieves nearly the same accuracy (96.6% vs. 97.4%) with half the parameters (0.53 M vs. 1.09 M) and less than half the memory footprint (2.03 MB vs. 4.16 MB). This equivalence in predictive quality at substantially lower computational cost underscores the effectiveness of the design choices refined through the ablation study, namely the adoption of Swish activations, depthwise-separable convolutions, and channel-attention modules.

High-capacity models such as EfficientNet-V2S/M/L and Xception reached 97–97.6% accuracy but required 20–118 million parameters, limiting their feasibility for real-time agricultural monitoring. Conversely, deeper residual and densely connected networks (ResNet50/101, DenseNet169) exhibited diminishing accuracy ($<95\%$) despite their large parameter counts. This pattern highlights that, in leaf disease recognition tasks, focused representational efficiency and well-balanced feature extraction outperform excessive architectural depth.

Table 5. Comparative performance of Light-MobileBerryNet and baseline models on the test set (values in bold correspond to the proposed model). The MACs and FLOPs are per image; latency corresponds to per-image inference time.

Model	Params (M)	Size (MB)	MACs (M)	FLOPs (M)	FPS	Latency (ms)	Accuracy	Precision	Recall	F1-Score	MCC
EfficientNet-V2L	118.08	450.44	12,309.34	24,618.68	≈6.5	≈153.8	0.976	0.977	0.976	0.976	0.967
EfficientNet-V2M	53.48	204.03	5406.72	10,813.44	≈12.3	≈81.3	0.975	0.976	0.975	0.975	0.966
EfficientNet-V2S	20.67	78.83	2877.19	5754.38	≈18.9	≈52.9	0.974	0.975	0.974	0.974	0.964
MobileNetV3-Small	1.09	4.16	59.60	119.21	205.29	4.87	0.9742	0.9754	0.9742	0.9744	0.970
Xception	21.39	81.62	4569.02	9138.04	≈13.3	≈75.0	0.973	0.974	0.972	0.972	0.962
EfficientNet-B3	11.18	42.66	992.99	1985.98	219.21	30.02	0.969	0.969	0.969	0.969	0.964
EfficientNet-B0	4.38	16.72	401.46	802.93	215.98	25.76	0.968	0.968	0.968	0.965	0.963
EfficientNet-B7	64.76	247.06	5264.81	10,529.62	≈11.5	≈86.9	0.970	0.977	0.969	0.969	0.962
Light-MobileBerryNet	0.53	2.03	143.97	287.94	131.85	7.58	0.966	0.966	0.967	0.969	0.960
DenseNet121	7.30	27.87	2851.11	5702.23	≈20.1	≈49.7	0.957	0.959	0.956	0.957	0.943
NASNetMobile	4.55	17.34	573.88	1147.77	≈85.5	≈11.7	0.952	0.955	0.952	0.953	0.939
DenseNet169	13.08	49.88	3380.19	6760.38	≈17.0	≈58.8	0.949	0.951	0.949	0.949	0.935
ResNet50	24.12	92.02	3877.57	7755.15	≈14.8	≈67.6	0.941	0.945	0.941	0.941	0.921
MobileNetV2	2.59	9.89	307.59	615.19	210.31	13.02	0.937	0.940	0.937	0.937	0.926
ResNet101	43.19	164.76	7599.18	15,198.36	≈9.8	≈101.9	0.927	0.930	0.926	0.926	0.905
MobileNetV3-Large	3.25	12.39	224.25	448.50	124.56	8.03	0.925	0.930	0.925	0.925	0.913

Figure 6 illustrates the accuracy–complexity relationship across models: while larger architectures achieve slightly higher scores, their computational requirements grow disproportionately. Beyond a moderate scale, increases in parameter count yield only marginal accuracy gains at a steep cost in latency and memory. Light-MobileBerryNet lies at the efficient end of this spectrum, delivering near–state-of-the-art accuracy with the smallest footprint and fastest inference. This balance of precision and efficiency highlights its suitability for edge-level agricultural deployment, where real-time decision support depends on both accuracy and resource economy.

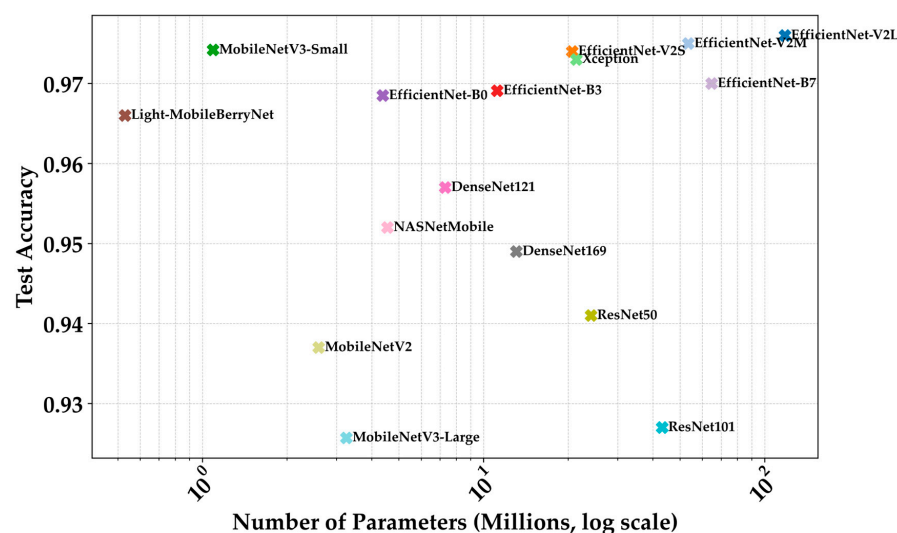


Figure 6. Accuracy–complexity trade-off across models. Light-MobileBerryNet achieves competitive accuracy with the lowest parameter count.

4.5. Visual Saliency Maps for Model Explainability

To provide deeper insight into the internal reasoning of the proposed Light-MobileBerryNet, we employed Grad-CAM to visualize class-discriminative regions across

the seven disease categories (Figure 7). The activation intensity ranges from blue (low) to red (high), denoting the spatial contribution of specific leaf or fruit areas to the network’s final decision. For angular leafspot and leaf spot, high activations concentrate along necrotic margins and chlorotic halos, while for anthracnose fruit rot and gray mold, the model attends to central decayed tissues and sporulating lesions on the fruit surface. In powdery mildew leaf and powdery mildew fruit, activations are localized over the whitish mycelial growth regions, indicating the network’s ability to distinguish superficial infections from background texture. Blossom blight responses align closely with petal necrosis and proximal leaf chlorosis, evidencing generalization to heterogeneous floral structures.

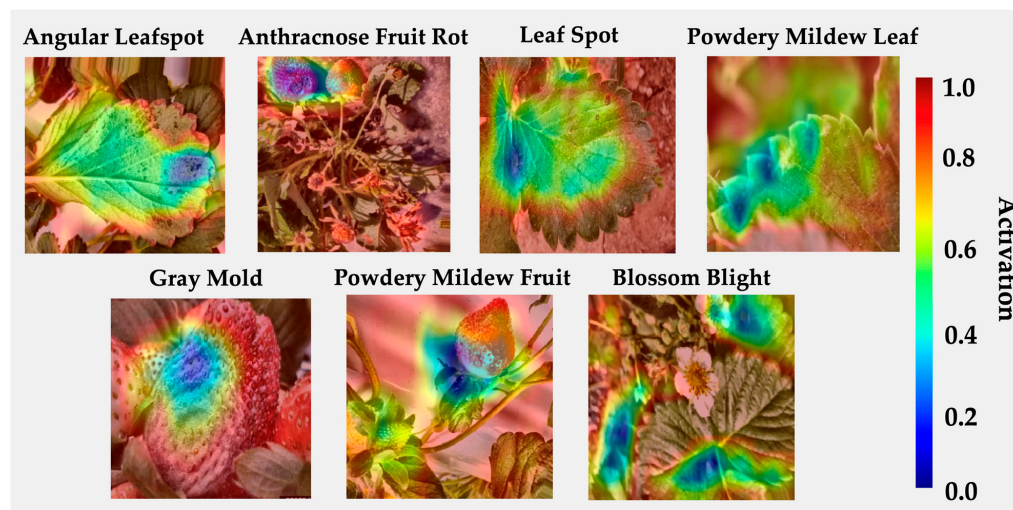


Figure 7. Grad-CAM visualizations for each disease category. Heatmaps highlight the symptomatic regions driving model predictions, confirming alignment with biologically relevant features.

To validate these qualitative observations, a quantitative correspondence analysis was conducted between Grad-CAM heatmaps and expert-annotated disease regions from Afzaal et al. work [32]. The dataset provides expert-defined infection severity levels—Level 1 (low–mid infection) and Level 2 (high infection)—derived from lesion spread, visual severity, and tissue maturity, with JSON annotations enabling direct overlap metrics. CAMs were normalized to $[0, 1]$ and binarized using a fixed per-image threshold τ ; heatmaps were resized to the annotation resolution before overlap computations. The resulting quantitative scores are summarized in Table 6, which reports the Intersection-over-Union (IoU), Dice coefficient, Pointing Game accuracy, and Energy In metrics for two expert-defined infection severity levels:

$$\text{IoU} = \frac{|A_p \cap A_g|}{|A_p \cup A_g|} \quad (22)$$

$$\text{Dice} = \frac{2|A_p \cap A_g|}{|A_p| + |A_g|} \quad (23)$$

$$\text{Pointing Game} = \frac{N_{hits}}{N_{total}} \quad (24)$$

$$E_{in}(\%) = \frac{\sum_{i \in A_g} CAM_i}{\sum_i CAM_i} \times 100 \quad (25)$$

where A_p and A_g denote the binarized Grad-CAM region and the expert lesion mask, respectively; N_{hits} is the number of images whose maximum CAM activation lies inside the expert mask; N_{total} is the number of evaluated images; and CAM_i is the activation intensity at pixel i .

Table 6. Quantitative interpretability metrics (IoU, Dice, Pointing Game, and Energy In) of Grad-CAM maps for Light-MobileBerryNet under two expert-defined infection severity levels.

Level	Infection Severity	IoU (Mean)	Dice (Mean)	Pointing Game (%)	Energy In (%)	IoU $\geq$ 0.30 (%)	IoU $\geq$ 0.50 (%)	Images
Level 1	Low-Mid	0.366	0.502	58.25	47.34	62.14	30.58	206
Level 2	High	0.320	0.432	56.42	44.63	54.75	31.66	537

These results support a consistent alignment between attention regions and expert-identified lesions, with slightly higher localization for mild infections (level 1). The moderate decline at level 2 is expected under increased occlusion, color blending, and lesion coalescence at advanced stages. Pointing Game > 56% indicates that peak Grad-CAM activations frequently coincide with the pathological core of the lesion, reinforcing the claim that the model focuses on biologically meaningful regions.

5. Discussion

Our findings demonstrate that high-performing recognition of strawberry leaf diseases can be achieved with architectures that are orders of magnitude lighter than conventional deep networks. This challenges the prevailing assumption that near-perfect accuracy in plant pathology inevitably requires large-scale models with tens of millions of parameters. By delivering 96.6% accuracy with fewer than one million parameters, Light-MobileBerryNet shows that compact, interpretable CNNs can approach state-of-the-art performance while remaining deployable in the low-power, offline environments where growers most urgently need them [1,9,10,19].

The comparative analysis underscores this trade-off. Architectures such as MobileNetV3-Small and EfficientNet-V2L achieved excellent classification in the test set, yet their size and computational demands restrict their real-world applicability. In contrast, Light-MobileBerryNet compresses this functionality into just 2 MB, reducing the barrier for mobile and edge implementation. These results echo recent reports in plant phenotyping and strawberry pathology, where deeper CNNs set the performance ceiling but also highlight the growing demand for lightweight, field-ready alternatives [10,17].

Interpretability adds another dimension of value. Grad-CAM maps confirmed that the proposed model consistently attended to biologically meaningful regions such as necrotic lesions in gray mold, powdery textures of mildew, and blight-affected petals, rather than background noise [10,26]. This capacity for visual explanation not only supports scientific validation but also fosters user trust, which is critical if models are to be adopted by farmers with limited technical training. Compared with hyperspectral pipelines that rely on specialized sensors, our RGB-based approach provides a lower-cost and more accessible route to disease monitoring without sacrificing transparency [27].

At the same time, certain limitations must be acknowledged. The relatively lower recall for powdery mildew fruit reflects the difficulty of detecting subtle texture cues that overlap with natural leaf structures [17,32]. Although the expanded seven-class dataset increases variability compared with the previous three-class configuration, it still does not fully represent the complexity of open-field conditions. In real agricultural environments, performance may decline due to uncontrolled illumination, partial occlusions from overlapping leaves or fruits, soil interference, and camera noise [48]. These factors have been shown to degrade classification accuracy in related studies and highlight the need for broader, more heterogeneous datasets and field validation campaigns that capture such variability. Further testing under real farm conditions will be essential to verify model robustness and adaptability beyond controlled greenhouse imagery [1,10,20,33].

Looking forward, several research directions emerge. First, the demonstrated efficiency makes Light-MobileBerryNet a strong candidate for offline mobile applications, allowing

farmers to diagnose diseases directly in the field without reliance on connectivity. Second, expanding the taxonomy beyond the seven target classes to include angular leaf scorch, additional powdery mildew forms (cenicilla), blossom blight, leaf scorch, and healthy leaves would increase robustness and practical value [1,5,18]. Third, further optimization through pruning and post-training quantization could compress the model even further, enabling faster inference on low-cost smartphones or embedded systems, for example, TinyML [45,49,50]. Integrating real-world data acquisition campaigns under variable environmental and lighting conditions would strengthen external validity and guide future model calibration for field deployment.

Finally, future work could explore multimodal extensions that combine conventional RGB imaging with complementary modalities such as thermal or hyperspectral data. Integrating additional spectral or temperature information may enhance the model's ability to detect early-stage infections or stress conditions that are not visually apparent in RGB imagery alone [14,27]. Such multimodal fusion approaches could further improve robustness under complex environmental conditions and support earlier, more reliable disease diagnosis in precision agriculture.

In summary, Light-MobileBerryNet provides a concrete demonstration that accurate, interpretable, and computationally efficient models are possible for strawberry leaf disease recognition. While heavier architectures still define the absolute upper bound of performance, the proposed approach offers a pragmatic balance between accuracy and deployability, bridging the gap between laboratory benchmarks and field-ready tools for precision agriculture.

6. Conclusions

Light-MobileBerryNet demonstrates that high accuracy and interpretability in strawberry disease recognition can be achieved without resorting to heavy, resource-intensive architectures. With only 0.53 million parameters and a 2 MB footprint, the model achieves a 96.6% accuracy while maintaining transparency through visual interpretability (Grad-CAM), proving that compact CNNs can remain both reliable and practical for real-world agricultural use.

Beyond its strong classification performance, Light-MobileBerryNet contributes three key advances: (1) it provides an efficient architecture optimized for mobile and edge deployment; (2) it balances accuracy, inference speed, and model size, reducing computational barriers for adoption in low-resource environments; and (3) it demonstrates that interpretable lightweight models can deliver field-relevant insights by highlighting biologically meaningful lesion regions.

Future work will focus on expanding the dataset to include additional diseases and healthy leaf samples, conducting field validations under variable environmental and lighting conditions, and implementing the model on mobile and embedded hardware platforms using TinyML or TensorFlow Lite to assess real-time performance and energy efficiency. These steps will help bridge the gap between laboratory research and scalable, farmer-accessible tools for precision agriculture.

Author Contributions: Conceptualization, R.O.-O. and A.G.-O.; methodology, R.O.-O., A.G.-O. and L.T.; software, A.G.-O., R.O.-O. and D.F.-D.; validation, R.O.-O., A.G.-O., K.L.P.-N. and L.T.; formal analysis, R.O.-O., A.G.-O. and A.Y.R.G.; investigation, R.O.-O., A.G.-O., D.F.-D. and A.Y.R.G.; resources, R.O.-O., L.T., A.Y.R.G. and D.F.-D.; data curation, A.G.-O., R.O.-O. and K.L.P.-N.; writing—original draft preparation, R.O.-O., A.G.-O. and A.Y.R.G.; writing—review and editing, R.O.-O., A.G.-O., L.T. and K.L.P.-N.; visualization, A.G.-O., R.O.-O. and D.F.-D.; supervision, R.O.-O. and A.Y.R.G.; project administration, R.O.-O., A.G.-O. and L.T.; funding acquisition, R.O.-O., D.F.-D. and L.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Tecnológico Nacional de México (TecNM), grant number 22298.25-P.

Data Availability Statement: The dataset used in this study, Strawberry Disease Detection Dataset, is publicly available on Kaggle at <https://www.kaggle.com/datasets/usmanafzaal/strawberry-disease-detection-dataset> (Accessed on 30 August 2025).

Acknowledgments: This work was developed in the framework of the international network “Red Internacional de Control y Cómputo Aplicados” supported by the TecNM. Additionally, the fourth author acknowledges support from CONAHCYT (SECIHTI, Mexico).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

σ	Sigmoid activation function
β	Scaling parameter of the Swish activation
δ	ReLU activation function
λ	Regularization coefficient in the loss function
ARS	Agricultural Research Service
BN	Batch Normalization
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNNs	Convolutional Neural Networks
ECA	Efficient Channel Attention
FN	False negatives
FP	False positives
FLOPs	Floating-point operations
FPS	Frames per second
GAP	Global Average Pooling
GPU	Graphics Processing Unit
H, W	Spatial height and width of the feature map
IoU	Intersection-over-Union
IR	Inverted Residual
K	Kernel size of convolution (e.g., 3×3)
M	Number of input channels in a convolutional layer
MACs	Multiply–Accumulate Operations
MCC	Matthews Correlation Coefficient
MLP	Multilayer Perceptron
N	Number of output channels in a convolutional layer
PR	Precision–Recall
r	Reduction ratio in the SE (Squeeze-and-Excitation) bottleneck
RGB	Red–Green–Blue
ROC-AUC	Receiver Operating Characteristic–Area Under the Curve
SE	Squeeze-and-Excitation
t	Expansion factor in inverted-residual block
tM	Expanded number of channels after applying the factor t
TB	Terabyte
TN	True negatives
TP	True positives
W_1, W_2	Weight matrices in the SE attention mechanism
$y, \hat{y}$	Ground truth and predicted class probability vectors
x	Input activation or feature map
YOLO	You Only Look Once

References

1. Karki, S.; Basak, J.K.; Tamrakar, N.; Deb, N.C.; Paudel, B.; Kook, J.H.; Kim, H.T. Strawberry disease detection using transfer learning of deep convolutional neural networks. *Sci. Hortic.* **2024**, *332*, 113241. [\[CrossRef\]](#)
2. Saha, R.; Shaikh, A.; Tarafdar, A.; Majumder, P.; Baidya, A.; Bera, U.K. Deep learning-based comparative study on multi-disease identification in strawberries. In Proceedings of the 2024 IEEE Silchar Subsection Conference (SILCON), Silchar, India, 15–17 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
3. Venkatesh, R.; Vijayalakshmi, K.; Geetha, M.; Bhuvanesh, A. Optimized deep belief network for multi-disease classification and severity assessment in strawberries. *J. Anim. Plant Sci.* **2025**, *35*, 482–497. [\[CrossRef\]](#)
4. Jiang, H.; Xue, Z.P.; Guo, Y. Research on plant leaf disease identification based on transfer learning algorithm. *J. Phys. Conf. Ser.* **2020**, *1576*, 012023. [\[CrossRef\]](#)
5. Parameshwari, V.; Brundha, A.; Gomathi, P.; Gopika, R. An intelligent plant leaf syndrome identification derived from pathogen-based deep learning algorithm by interfacing IoT in smart irrigation system. In Proceedings of the 2nd Int. Conf. on Artificial Intelligence and Machine Learning Applications (AIMLA), Tiruchengode, India, 15–16 March 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–7.
6. Kaushik, A.; Attri, S.H.; Chauhan, S.S. Elucidating deep transfer learning approach for early plant disease detection through spot and lesion analysis. In Proceedings of the 3rd International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 7–8 March 2025; IEEE: Piscataway, NJ, USA, 2025; pp. 1145–1150.
7. Adiga, A.; Gagandeep, N.K.; Prabhu, A.A.; Pai, H.; Kumar, R.A. Comparative analysis on deep learning models for plant disease detection. In Proceedings of the International Conference on Emerging Technologies in Computer Science for Interdisciplinary Applications (ICETCS), Bengaluru, India, 22–23 April 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
8. Wang, W.; Chen, W.; Xu, D.; An, Y. Diseased plant leaves identification by deep transfer learning. In *Artificial Intelligence Technologies and Applications*; Chen, C., Ed.; IOS Press: Amsterdam, The Netherlands, 2024.
9. Aybergüler, A.; Arslan, E.; Kayaarma, S.Y. Deep learning-based growth analysis and disease detection in strawberry cultivation. In Proceedings of the 2023 Innovations in Intelligent Systems and Applications Conference (ASYU), Sivas, Turkey, 11–13 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5.
10. Wang, J.; Li, Z.; Gao, G.; Wang, Y.; Zhao, C.; Bai, H.; Li, Q. BerryNet-lite: A lightweight convolutional neural network for strawberry disease identification. *Agriculture* **2024**, *14*, 665. [\[CrossRef\]](#)
11. Shang, C.; Wu, F.; Wang, M.; Gao, Q. Cattle behavior recognition based on feature fusion under a dual attention mechanism. *J. Vis. Commun. Image Represent.* **2022**, *85*, 103524. [\[CrossRef\]](#)
12. Singh, R.; Sharma, N.; Gupta, R. Strawberry leaf disease detection using transfer learning models. In Proceedings of the IEEE 2nd Int. Conf. on Industrial Electronics: Developments & Applications (ICIDEA), Imphal, India, 29–30 September 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 341–346.
13. Hu, X.; Xu, T.; Wang, C.; Zhu, H.; Gan, L. Domain generalization method of strawberry disease recognition based on instance whitening and restitution. *Smart Agric.* **2025**, *7*, 124–135.
14. Jiang, Q.; Wu, G.; Tian, C.; Li, N.; Yang, H.; Bai, Y.; Zhang, B. Hyperspectral imaging for early identification of strawberry leaves diseases with machine learning and spectral fingerprint features. *Infrared Phys. Technol.* **2021**, *118*, 103898. [\[CrossRef\]](#)
15. Tumpa, S.B.; Halder, K.K. A comparative study on different transfer learning approaches for identification of plant diseases. In Proceedings of the Int. Conf. on Next-Generation Computing, IoT and Machine Learning (NCIM), Gazipur, Bangladesh, 16–17 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
16. Das, S.; Karna, H.B.; Das, S.; Hazra, R. Disease detection in plant leaves using transfer learning. In Proceedings of the First Int. Conf. on Electronics, Communication and Signal Processing (ICECSP), New Delhi, India, 8–10 August 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
17. Wang, J.; Li, J.; Meng, F. Recognition of strawberry powdery mildew in complex backgrounds: A comparative study of deep learning models. *AgriEngineering* **2025**, *7*, 182. [\[CrossRef\]](#)
18. Pertiwi, S.; Wibowo, D.H.; Widodo, S. Deep learning model for identification of diseases on strawberry (*Fragaria* sp.) plants. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2023**, *13*, 1342–1348. [\[CrossRef\]](#)
19. Dong, C.; Zhang, Z.; Yue, J.; Zhou, L. Automatic recognition of strawberry diseases and pests using convolutional neural network. *Smart Agric. Technol.* **2021**, *1*, 100009. [\[CrossRef\]](#)
20. Shin, J.; Chang, Y.K.; Heung, B.; Nguyen-Quang, T.; Price, G.W.; Al-Mallahi, A. A deep learning approach for RGB image-based powdery mildew disease detection on strawberry leaves. *Comput. Electron. Agric.* **2021**, *183*, 106042. [\[CrossRef\]](#)
21. Li, G.; Jiao, L.; Chen, P.; Liu, K.; Wang, R.; Dong, S.; Kang, C. Spatial Convolutional Self-Attention-Based Transformer Module for Strawberry Disease Identification under Complex Background. *Comput. Electron. Agric.* **2023**, *212*, 108121. [\[CrossRef\]](#)
22. Aghamohammadesmaeilketabforoosh, K.; Nikan, S.; Antonini, G.; Pearce, J.M. Optimizing Strawberry Disease and Quality Detection with Vision Transformers and Attention-Based Convolutional Neural Networks. *Foods* **2024**, *13*, 1869. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Nguyen, H.T.; Tran, T.D.; Nguyen, T.T.; Pham, N.M.; Nguyen Ly, P.H.; Luong, H.H. Strawberry Disease Identification with Vision Transformer-Based Models. *Multimed. Tools Appl.* **2024**, *83*, 73101–73126. [\[CrossRef\]](#)
24. Chen, M.; Zou, W.; Niu, X.; Fan, P.; Liu, H.; Li, C.; Zhai, C. Improved YOLOv8-based segmentation method for strawberry leaf and powdery mildew lesions in natural backgrounds. *Agronomy* **2025**, *15*, 525. [\[CrossRef\]](#)
25. Alhwaiti, Y.; Khan, M.; Asim, M.; Siddiqi, M.H.; Ishaq, M.; Alruwaili, M. Leveraging YOLO deep learning models to enhance plant disease identification. *Sci. Rep.* **2025**, *15*, 7969. [\[CrossRef\]](#)
26. Kim, H.; Kim, D. Deep-learning-based strawberry leaf pest classification for sustainable smart farms. *Sustainability* **2023**, *15*, 7931. [\[CrossRef\]](#)
27. Ou, Y.; Yan, J.; Liang, Z.; Zhang, B. Hyperspectral imaging combined with deep learning for the early detection of strawberry leaf gray mold disease. *Agronomy* **2024**, *14*, 2694. [\[CrossRef\]](#)
28. Xu, M.; Yoon, S.; Jeong, Y.; Park, D.S. Transfer learning for versatile plant disease recognition with limited data. *Front. Plant Sci.* **2022**, *13*, 1010981. [\[CrossRef\]](#)
29. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Uribe-Toscano, S. A robust transfer learning approach with histopathological images for lung and colon cancer detection using EfficientNetB3. *Healthc. Anal.* **2025**, *7*, 100391. [\[CrossRef\]](#)
30. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Uribe-Toscano, S. Enhancing early lung cancer detection with MobileNet: A comprehensive transfer learning approach. *Franklin Open* **2025**, *10*, 100222. [\[CrossRef\]](#)
31. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; García-Rodríguez, J.A. A hybrid deep learning and machine learning approach with Mobile-EfficientNet and Grey Wolf Optimizer for lung and colon cancer histopathology classification. *Cancers* **2024**, *16*, 3791. [\[CrossRef\]](#)
32. Afzaal, U.; Bhattarai, B.; Pandeya, Y.R.; Lee, J. An instance segmentation model for strawberry diseases based on Mask R-CNN. *Sensors* **2021**, *21*, 6565. [\[CrossRef\]](#)
33. Narla, V.L.; Suresh, G.; Rao, C.S.; Awadh, M.A.; Hasan, N. A multimodal approach with firefly-based CLAHE and multiscale fusion for enhancing underwater images. *Sci. Rep.* **2024**, *14*, 27588. [\[CrossRef\]](#)
34. Islam, T.; Hafiz, M.S.; Jim, J.R.; Kabir, M.M.; Mridha, M.F. A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions. *Healthc. Anal.* **2024**, *5*, 100340. [\[CrossRef\]](#)
35. Gao, X.; Xiao, Z.; Deng, Z. High accuracy food image classification via vision transformer with data augmentation and feature augmentation. *J. Food Eng.* **2024**, *365*, 111833. [\[CrossRef\]](#)
36. Sunkari, S.; Sangam, A.; Raman, R.; Rajalakshmi, R. A refined ResNet18 architecture with Swish activation function for diabetic retinopathy classification. *Biomed. Signal Process. Control* **2024**, *88*, 105630. [\[CrossRef\]](#)
37. Javanmardi, S.; Ashtiani, S.H.M. AI-Driven Deep Learning Framework for Shelf Life Prediction of Edible Mushrooms. *Postharvest Biol. Technol.* **2025**, *222*, 113396. [\[CrossRef\]](#)
38. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for Activation Functions. *arXiv* **2017**, arXiv:1710.05941. [\[CrossRef\]](#)
39. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2024; pp. 7132–7141.
40. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861. [\[CrossRef\]](#)
41. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2024; pp. 4510–4520.
42. Tummala, S.; Kadry, S.; Nadeem, A.; Rauf, H.T.; Gul, N. An explainable classification method based on complex scaling in histopathology images for lung and colon cancer. *Diagnostics* **2023**, *13*, 1594. [\[CrossRef\]](#)
43. Nobel, S.N.; Afroj, M.; Kabir, M.M.; Mridha, M.F. Development of a cutting-edge ensemble pipeline for rapid and accurate diagnosis of plant leaf diseases. *Artif. Intell. Agric.* **2024**, *14*, 56–72. [\[CrossRef\]](#)
44. Mittapalli, P.S.; Tagore, M.R.N.; Reddy, P.A.; Kande, G.B.; Reddy, Y.M. Deep Learning-Based Real-Time Object Detection on Jetson Nano Embedded GPU. In *Microelectronics, Circuits and Systems: Select Proceedings of Micro2021*; Springer Nature: Singapore, 2023; pp. 511–521.
45. Mayo, D.; Cummings, J.; Lin, X.; Gutfreund, D.; Katz, B.; Barbu, A. How Hard Are Computer Vision Datasets? Calibrating Dataset Difficulty to Viewing Time. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 11008–11036.
46. Ochoa-Ornelas, R.; Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Uribe-Toscano, S. Lung and colon cancer detection with Inception-ResNetV2: A transfer learning approach. *J. Res. Dev.* **2024**, *10*, e11025113. [\[CrossRef\]](#)
47. Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Ochoa-Ornelas, R.; Ruiz-Velazquez, E.; Uribe-Toscano, S.; Cuevas-Chávez, J.I.; Sánchez-Arias, D.A. Non-invasive multiclass diabetes classification using breath biomarkers and machine learning with explainable AI. *Diabetology* **2025**, *6*, 51. [\[CrossRef\]](#)
48. Tamrakar, N.; Paudel, B.; Karki, S.; Deb, N.C.; Arulmozhi, E.; Kook, J.H.; Kim, H.T. Peduncle Detection of Ripe Strawberry to Localize Picking Point Using DF-Mask R-CNN and Monocular Depth. *IEEE Access* **2025**, *13*, 1–12. [\[CrossRef\]](#)

49. Gudiño-Ochoa, A.; García-Rodríguez, J.A.; Ochoa-Ornelas, R.; Cuevas-Chávez, J.I.; Sánchez-Arias, D.A. Noninvasive Diabetes Detection through Human Breath Using TinyML-Powered E-Nose. *Sensors* **2024**, *24*, 1294. [[CrossRef](#)] [[PubMed](#)]
50. Samanta, R.; Saha, B.; Ghosh, S.K. TinyML-on-the-Fly: Real-Time Low-Power and Low-Cost MCU-Embedded On-Device Computer Vision for Aerial Image Classification. In Proceedings of the 2024 IEEE Space, Aerospace and Defence Conference (SPACE), Oxford, UK, 8–10 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 194–198.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.