ORIGINAL ARTICLE

Detection of diseases in cucumber using deep neural networks

Andrea Menco-Tovar · Juan Carlos Martinez-Santos · Edwin Puertas

Received: 23 April 2025 / Accepted: 12 January 2026 / Published online: 6 March 2026
© The Author(s) 2026

Abstract
The cucumber (*Cucumis sativus* L.), a globally essential crop, faces severe threats from various foliar diseases. This work explores deep neural networks (AlexNet, Vision Transformer, MobileNet, and U-Net) for the early and accurate detection of these pathologies based on leaf images. We analyzed 4,353 images classified as healthy or diseased through advanced preprocessing and data augmentation techniques. The results highlight Vision Transformer as the most effective architecture, achieving 99% accuracy, surpassing MobileNet with similar performance. Meanwhile, AlexNet and U-Net demonstrated more limited performance. The research underscores the practical applicability of these technologies in intelligent agriculture systems, promoting informed decision-making to reduce economic losses and environmental impact. Furthermore, it emphasizes the importance of integrating these tools into low-cost devices for implementation in rural areas. This approach contributes to the sustainability of cucumber cultivation. It sets a precedent for the efficient management of diseases in modern agriculture.

Keywords Disease detection · Computer vision · Deep learning · Convolutional Neural Networks (CNN) · Crops

1 Introduction

Agriculture is a fundamental pillar of global food security, underpinning the economy and human well-being. Authors estimated that more than 85% of the population depends directly or indirectly on this activity, which generates food, fiber, and other products essential for modern life [1–3]. However, projected population growth and adverse climatic conditions pose critical challenges to meeting the growing global demand for food, which will increase by 50% by 2050 [4–6]. Among the main challenges in agriculture is the timely detection and management of foliar diseases, which severely affect crop yields and quality. The appearance of leaf spots, discolorations, and lesions are common signs of infections caused by bacteria, fungi, or viruses [7–10]. Early identification of these pathologies, especially in extensive agricultural environments, requires efficient and accurate techniques that overcome the limitations of manual methods such as visual inspection [11–13].

Cucumber (*Cucumis sativus* L.) is one of the most important vegetable crops globally, not only for its economic relevance but also for its nutritional value and versatility. This vegetable, belonging to the family Cucurbitaceae,

is widely cultivated in tropical and subtropical regions due to its adaptability to hot and humid climates [14–16]. Cucumber fruits are consumed fresh in salads, pickled or cooked, noted for their high water content (93-96%) and their supply of essential nutrients such as vitamins C and K, potassium, and antioxidants [17]. In addition, cucumber cultivation is a significant source of income in several agricultural economies, especially in Asia and Europe, where it occupies a substantial share of protected crop areas [18].

Despite its benefits, cucumber is highly susceptible to a wide range of diseases, including fungal infections such as downy mildew *(Pseudoperonospora cubensis)* and powdery mildew *(Erysiphe cichoracearum)*, which affect both fruit quality and yield [18–20]. These diseases reduce yields in the field and increase production costs due to using fungicides and other control measures. Early and accurate detection of these pathologies is crucial to minimize economic losses and ensure crop sustainability [21–23]. Furthermore, its sensitivity to environmental factors, such as excessive moisture or temperature extremes, makes it an ideal model to study the interaction between pathogens and crops under stress [24–27].

Despite advances in crop disease detection using deep learning techniques, challenges remain in the early and accurate identification of cucumber pathologies. Traditional methods based on visual inspection are subjective and require specialized knowledge, while approaches based on artificial intelligence face limitations related to the variability of environmental conditions, image quality, and model generalization capability. In this context, it is necessary to evaluate different deep learning architectures to determine which performs better in detecting cucumber leaf disease, ensuring an efficient and adaptable solution to field conditions.

In this context, deep learning architectures, particularly convolutional neural networks (CNN), have emerged as key tools for real-time foliar disease detection. These technologies overcome the limitations of traditional methods by automating the detection process and learning directly from the data without the need for manual feature design [28–31]. One of the most notable advances of these architectures is their accuracy and robustness. By employing millions of adjustable parameters, CNN can capture complex patterns in the data, even in challenging lighting conditions and backgrounds. It makes them ideal for agricultural tasks where the collected images may contain significant variations in context and quality [32–34]. In addition, their generalisability sets them apart from traditional methods, which often fail in the face of new variations in disease or growing conditions. CNNs, when trained with diverse data, can better adapt to unseen scenarios, increasing their practical utility in field environments [35, 36].

Real-time efficiency is another key advantage of modern architectures. Models such as Vision Transformer are optimized for rapid inference and enable real-time disease detection, even on resource-constrained devices such as agricultural drones or smart cameras. This capability transforms the way farmers monitor their crops, reducing response times to disease outbreaks and optimizing decision-making [31, 37, 38]. Finally, these architectures are deeply integrated into innovative farming systems, contributing to sustainability. CNNs enable informed decisions on water, fertilizers, and pesticides, reducing environmental impact and improving operational efficiency. By integrating with technologies such as IoT sensors and remote monitoring systems, these advanced tools enhance farmers' ability to manage their crops accurately and sustainably [35, 39].

The article's primary contribution lies in its focus on intelligent agriculture, specifically through a comparative evaluation of various deep neural network architectures (AlexNet, Vision Transformer, MobileNet, and U-Net) for the early and accurate detection of foliar diseases in cucumber plants. Moreover, the comparative analysis provides a practical guide to selecting more effective models in real agricultural contexts. Incorporating advanced data preprocessing and augmentation techniques strengthens the generalization capability of the models in real and varied situations. Additionally, we ensured the replicability of the models employed in this research by thoroughly and meticulously describing the dataset, preprocessing techniques, data augmentation, and the specific parameters utilized in training each neural architecture. The public availability of the data used is also beneficial in enabling other researchers to reproduce and validate these results in different agricultural contexts, thereby contributing to advancing and consolidating knowledge in deep learning applications in intelligent agriculture.

In our research, we combined two cucumber leaf image datasets available at[1] and.[2] We compared architectures such as AlexNet, Vision Transformer (ViT), MobileNet, and U-Net to detect and identify cucumber diseases early and reliably. The rest of this article is as follows: Sect. 2 presents state of the art. Section 3 includes the methodology broken down into Preprocessing, Dataset, and Technique. Section 4 presents the results of this proposed work. Finally, Sect. 5 concludes the paper.

2 Related work

Deep learning has become the dominant paradigm for plant disease classification, surpassing traditional machine learning methods such as SVM, random forests, and XGBoost due to its superior capability for hierarchical feature extraction and robust generalization. Classical CNN architectures–including VGG16, ResNet50, InceptionV3, Xception, and MobileNetV2–remain strong baselines; however, recent studies in cucumber disease recognition have advanced toward more specialized, high-efficiency, and explainable models.

Recent works illustrate significant progress in cucumber disease detection. Qiao et al. [15] enhanced YOLOv5s with transformer modules and a microscopic-object detection head for downy mildew spore identification, achieving 95.4% accuracy. Zhu et al. [24] introduced GCS-YOLOv8, a method tailored for detecting small fungal structures under challenging illumination, integrating XAI techniques to validate model robustness, and reaching 95.2% accuracy. Complementary advances include multimodal contrastive learning approaches for disease recognition Cao et al. [19], lightweight severity estimation pipelines optimized for MobileNetV3 Li et al. [40], and MobileNetV3-based training strategies using flooding loss to improve generalization in natural field conditions Liu et al. [26]. Additional research explores multi-scale feature fusion networks and attention-guided architectures for improved pixel-level discrimination Dai et al. [11], as well as the incorporation of explainability techniques such as Grad-CAM to interpret lesion regions and infection stages Hernandez et al. [28].

As these highly specialized models continue to evolve, a parallel research direction focuses on ensuring that such capabilities remain practical in real-world agricultural settings, where devices often have limited computational resources. Given the need for real-time diagnosis in resource-limited agricultural environments, recent literature emphasizes lightweight models optimized for mobile and embedded devices. Wang et al. [41] proposed an ultra-efficient CNN using depthwise separable convolutions to reduce parameters by over 80% without sacrificing accuracy. Duhan [42] introduced a compact attention-enhanced CNN achieving competitive accuracy with reduced inference time for smartphone deployment. Studies also demonstrate that quantization-aware training enables 8-bit models to maintain accuracy while supporting inference on low-cost devices such as Raspberry Pi and microcontrollers [43].

In addition to efficiency, another line of research seeks to more faithfully represent the complexity of plant pathology, since field conditions often involve overlapping or co-occurring symptoms. While most cucumber disease datasets are single-label, modeling real-world conditions requires addressing co-occurring symptoms. Aboelenin et al. [44] presented a CNN–ViT hybrid with hierarchical label encoding for multi-label plant disease prediction. Wei et al. [45] proposed a lightweight few-shot learning framework capable of recognizing multiple diseases and pests with minimal annotated samples. Recent reviews also highlight the growing relevance of multi-label datasets PlantVillage-ML, PD-ML for supporting complex field-level symptom interactions [46].

Extending beyond static image analysis, contemporary efforts also explore systems capable of continuous and context-aware monitoring, recognizing that disease detection increasingly occurs within broader precision agriculture pipelines. Beyond static leaf imagery, real-time monitoring has gained relevance through IoT-enabled devices, edge processing, and UAV platforms. Tariq et al. [47] proposed an integrated smart agriculture framework combining lightweight deep learning models with sensor data for real-time crop health surveillance.

[1] https://www.kaggle.com/datasets/kaushigihanml/cucumber-leaf-disease-dataset

[2] https://www.kaggle.com/code/akshayr009/cucumberdiseases/input

Sharafat et al. [48] developed an edge-only inference pipeline optimized for low-latency detection in continuous monitoring setups. From the aerial perspective, Kouadio et al. [49] reviewed UAV-based disease detection challenges, including motion blur, variable lighting, and large-scale field coverage.

Given the diversity of methods, modalities, and operating conditions across the literature, standardized datasets remain essential to enable fair benchmarking and to study generalization. Benchmark datasets such as PlantVillage,[3] IP102,[4] CVPPP,[5] and domain-specific cucumber leaf datasets play a crucial role in evaluating model performance under standardized conditions. Recent research stresses cross-domain evaluation testing models on both controlled and field images to quantify generalization gaps and dataset biases. The present study aligns with these practices by integrating two heterogeneous datasets with distinct acquisition settings and background complexities

3 Methods

This section outlines the methodological framework followed in the study. First, it describes the integration and harmonization of two heterogeneous cucumber disease datasets, including their restructuring into a unified healthy–sick scheme and their subsequent stratified division into training, validation, and test sets. The next part details the preprocessing procedures applied to standardize the images for each architecture, covering resizing, normalization, tailored data augmentation strategies, and the automatic mask-generation pipeline used to enable U-Net–based lesion localization. Finally, the section presents the deep-learning models and training procedures employed–AlexNet, MobileNet, ViT, and U-Net–summarizing their configuration, optimization strategies, and roles in both classification and segmentation tasks.

3.1 Dataset

The study integrates two publicly available datasets from Kaggle: Cucumber Leaf Disease Dataset and Cucumber Diseases, Table 1 presents each dataset's data. Since both datasets present different folder structures and disease class definitions, a harmonization process was conducted to obtain a unified binary classification scheme healthy vs sick. Fig. 1 presents some examples.

All diseased categories Powdery mildew, Downy mildew, III_cucumber were merged into a single sick class, while Healthy leaf and good_Cucumber were merged into healthy. No additional relabeling was required, and no samples were removed during harmonization.

A 70%, 15%, 15% stratified split into train, validation, and test sets was applied at the folder level with a fixed seed (42) to ensure reproducibility. No oversampling, undersampling, or class weighting was applied; the models

Table 1 Dataset of cucumber leaf diseases

Name	Category	Number of images	Image size
Cucumber Leaf Disease Dataset	Healthy leaf Powdery mildew Downy mildew	3754	768 x 768 pixel
Cucumber Diseases	Healthy Sick	695	1280 x 606 pixel

[3] https://github.com/tensorflow/datasets/blob/master/tensorflow_datasets/datasets/plant_village/plant_village_dataset_builder.py

[4] https://github.com/xpwu95/IP102?utm_source

[5] https://www.plant-phenotyping.org/CVPPP2014-dataset?utm_source

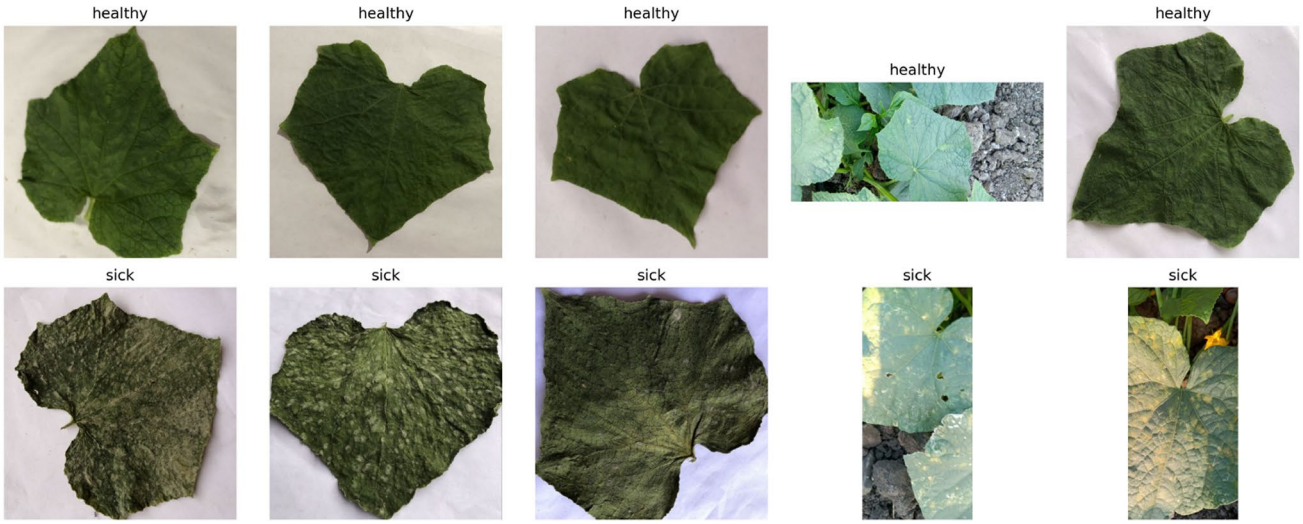


Fig. 1 Sample images from the cucumber leaf image dataset

Table 2 Input size and normalization applied per model

Model	Input size	Normalization
AlexNet	227 × 227	Rescale 1/255
MobileNet	224 × 224	Rescale 1/255
ViT	224 × 224	Mean = (0.5), Std = (0.5)
U-Net	128 × 128	Rescale 1/255

were trained using the natural class imbalance 2777 sick vs. 1576 healthy, as preliminary experiments with balanced datasets resulted in lower performance metrics.

3.2 Pre-processing

Images were standardized according to the input requirements of each architecture, see Table 2:

For data augmentation, each model was configured according to its architectural characteristics. AlexNet received moderate geometric transformations, including rotations of up to 20°, horizontal and vertical shifts, and horizontal flipping. MobileNet employed a broader augmentation regime with rotations of up to 40°, width–height shifts, shear transformations, zoom operations, and horizontal flipping. In contrast, ViT and U-Net did not use geometric augmentation; both models operated exclusively with resizing and standard normalization procedures. All augmentation outputs were inspected visually a common practice in plant pathology image analysis although no formal artifact-detection mechanism was implemented. Likewise, no explicit image-quality filtering was applied, and all images were used exactly as provided by the dataset authors.

Although the main task focused on classification, the study incorporated a U-Net segmentation model to explore lesion-level localization. Since neither dataset supplied segmentation masks, the authors implemented an automatic mask-generation pipeline. This process began by computing the 5th–95th percentile HSV statistics for all sick images, from which a global HSV range was derived. Using this range, masks were generated through the cv2.inRange function, whereas healthy images were assigned an all-zero mask. These automatically produced masks represent approximate lesion regions rather than expert annotations. Their adequacy was assessed both visually and quantitatively using pixel-level accuracy, recall, F1-score, and AUC.

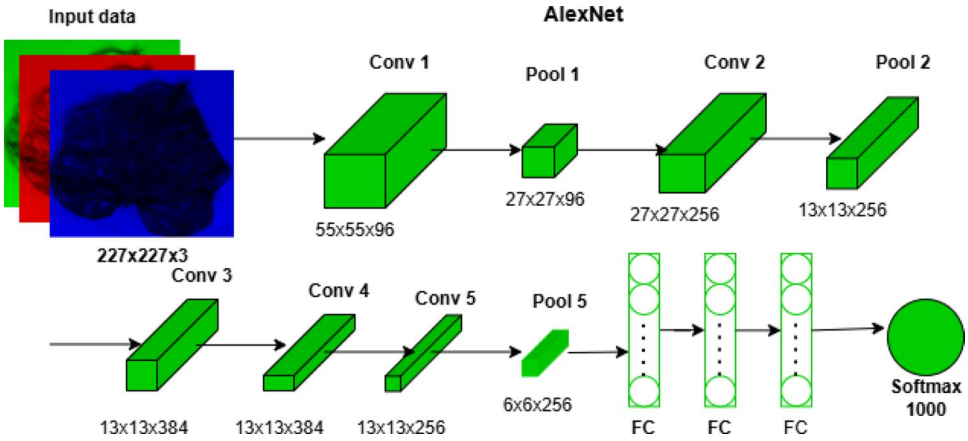


Fig. 2 Configuration of the architecture AlexNet

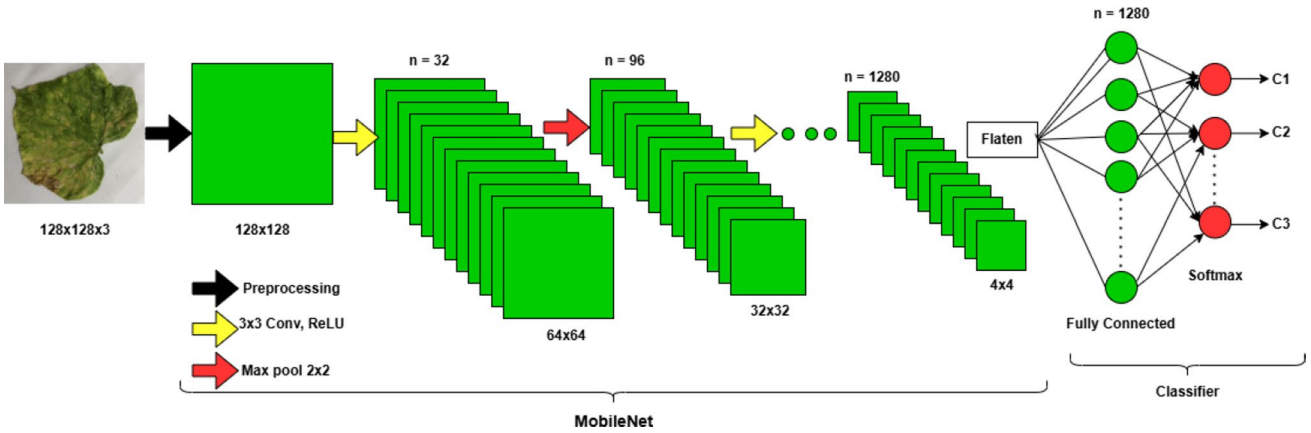


Fig. 3 Configuration of the architecture MobileNet

3.3 Technique

In recent years, deep neural network architectures have transformed the field of computer vision, offering innovative solutions for tasks such as object detection, image classification, and semantic segmentation. These architectures, designed in various configurations and approaches, have proven highly effective in multiple applications, from mobile devices to biomedical systems. This paper explores four representative architectures in this domain: AlexNet, Vision Transformer (ViT), MobileNet, and U-Net. Each of these architectures incorporates advanced block and layer design techniques, optimizing their performance for specific use cases. We described the main features of each model below, highlighting their structural innovations and their impact on the field.

3.3.1 AlexNet

AlexNet comprises five convolutional blocks and three fully connected layers, as shown in Fig. 2. Each convolutional block includes convolution layers with large filter sizes, followed by max pooling operations and ReLU activation functions. It also introduces Dropout layers in the final stage to mitigate overfitting and improve generalization. The design uses local normalization operations to accelerate convergence during training, optimizing GPU usage [50]. The original AlexNet architecture was adapted for disease detection in cucumber leaves. This adaptation involved adjusting the input layer to accept images resized to 224 × 224 pixels. Furthermore, we modified the final fully connected layers to perform binary classification (healthy or diseased leaves), simplifying the model output and thus optimizing training for this particular task.

3.3.2 MobileNet

MobileNet is composed of Depthwise Separable Convolution blocks, which divide convolutional operations into two stages: depthwise convolution (per channel) and pointwise convolution (1 × 1), as shown Fig. 3. These blocks include ReLU6 activation functions and BatchNorm normalization layers to improve stability during training. In addition, MobileNet v2 and later incorporate residual inverted blocks with shortcut connections, which will enhance efficiency by maintaining important features in low-dimensional spaces [51, 52]. MobileNet was adapted by leveraging its lightweight structure based on depth-separable convolutions. It is well-suited for devices with limited resources. The specific adaptation entailed adjusting the input dimension to 224x224 pixels and redefining the last convolutional and classification layers to optimize binary detection. It ensured high computational performance and accuracy in foliar disease classification.

3.3.3 Vision transformer (ViT)

The Vision Transformer (ViT) uses a structure based on traditional transformer blocks adapted to computer vision, as shown in Fig. 4. Each block includes Multi-Head Self-Attention (MHSA) and fully connected feed-forward layers. We divided the images into fixed patches processed as tokens. We added positional coding to retain spatial information. ViT employs techniques such as the use of LayerNorm normalization layers before self-attenuation operations and masking in pretextual training, allowing global relationships to be captured efficiently [40, 53].

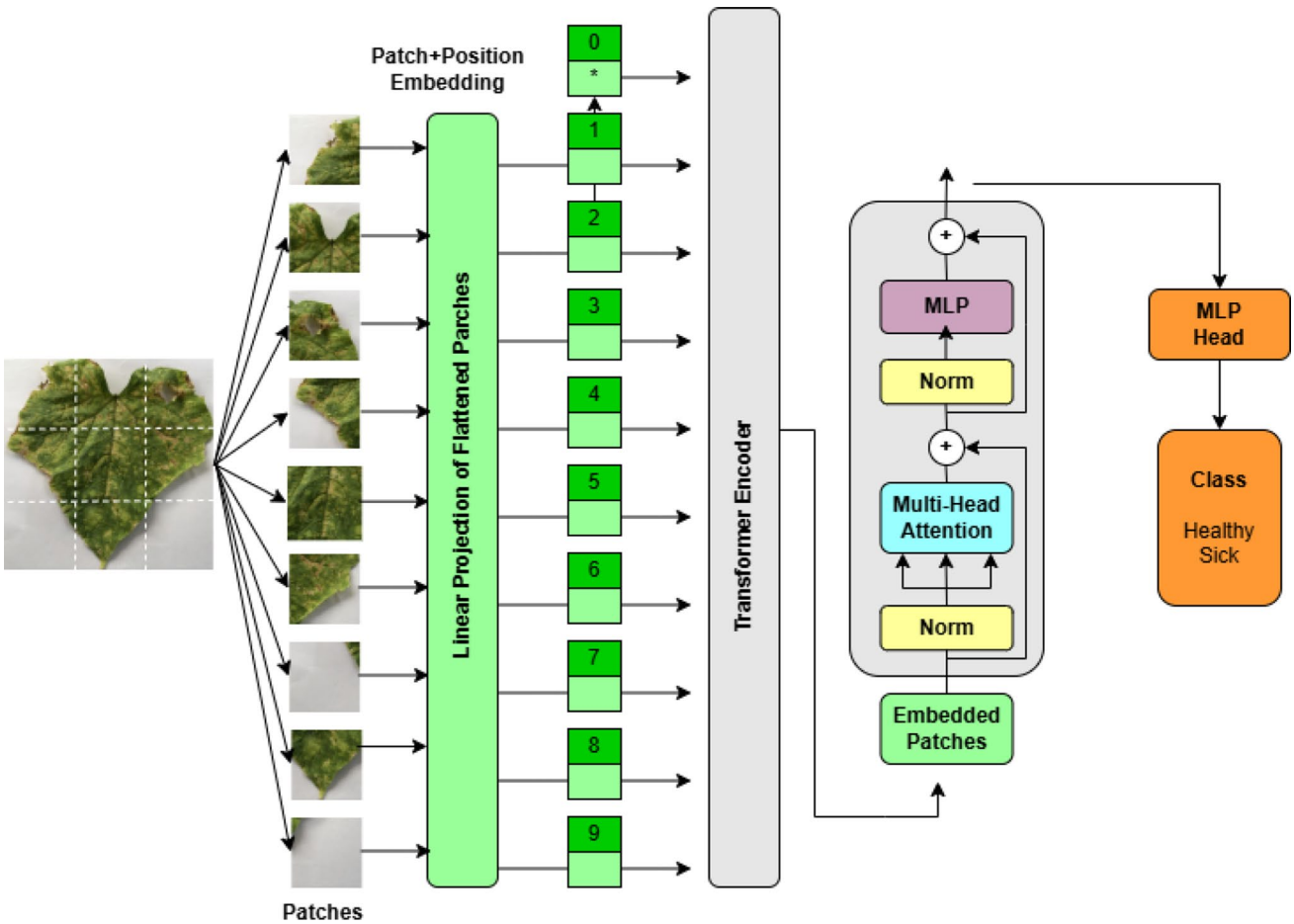


Fig. 4 Configuration of the architecture vision transformer (ViT).

For ViT, the adaptation entailed segmenting each 224x224 pixel image into fixed patches, which we then converted into tokens. Then, we processed these tokens through a self-attenuation mechanism. This process enabled the capture of significant global spatial relationships in cucumber leaf images. It, in turn, facilitated accurate disease identification from subtle patterns in leaf textures and colors.

3.3.4 U-Net

We organized U-Net in a symmetric design with two main paths: the encoder and the decoder. Fig. 5 shows the architecture. The encoder consists of convolutional blocks that apply two 3×3 convolutions followed by ReLU activations and max pooling. The decoder uses similar blocks but incorporates upsampling operations to reconstruct the spatial resolution. U-Net stands out for its skip connections, which transfer features from the encoder to the decoder, preserving fine details. Furthermore, convolutional layers at the end of each decoding stage improve the accuracy in edge segmentation [54, 55]. Although we principally designed U-Net for semantic segmentation, it was adapted for binary classification by adjusting the final decoder output. Rather than generating intricate segmentation masks, we integrated an auxiliary layer to condense the features extracted by the network into a straightforward classification, distinguishing between healthy and diseased leaves. This modification enabled the utilization of U-Net's capacity to discern intricate details. However, it concomitantly constrained its performance compared to architectures explicitly designed for classification tasks.

3.4 Model architectures and training procedures

Four representative deep-learning paradigms were evaluated in this study. The first model, AlexNet, was trained from scratch as a conventional CNN baseline. It used the Adam optimizer with a learning rate of approximately 1×10^{-3}, categorical cross-entropy as the loss function, and a batch size of 32 over 25 epochs. The fully connected layers incorporated dropout with a rate of 0.5 to reduce overfitting, and the model was trained using moderate geometric data augmentation. No early-stopping strategy was applied, and the training followed a single-run protocol.

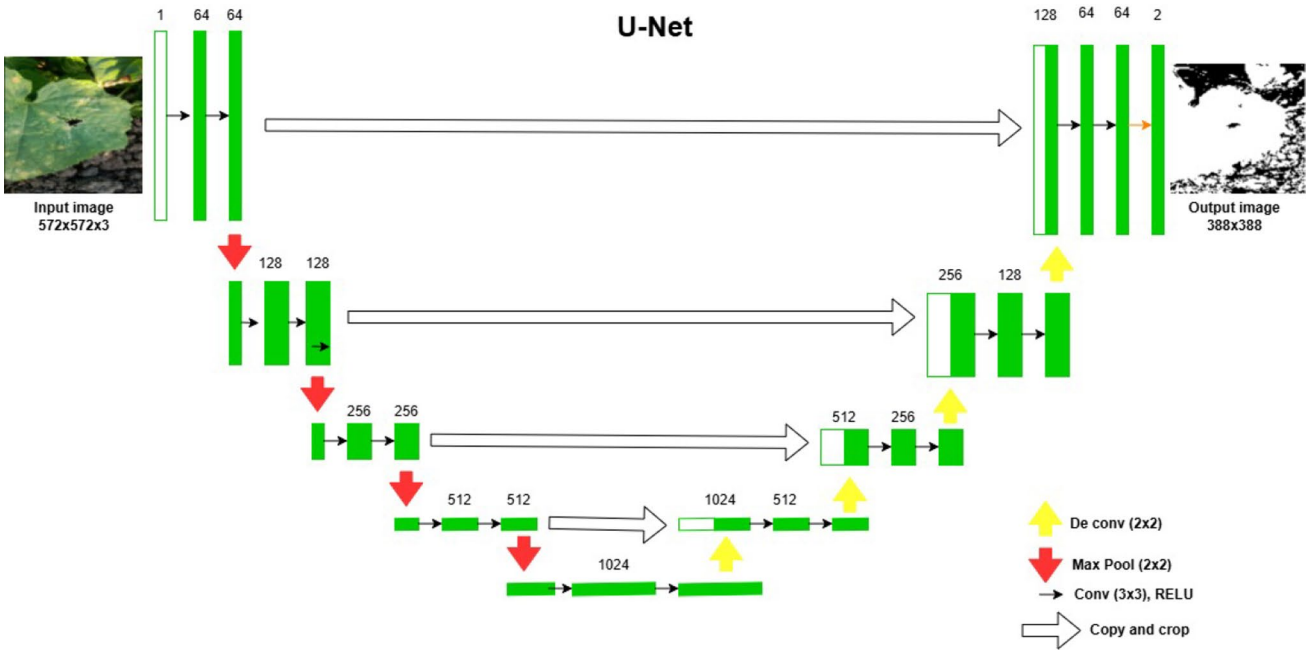


Fig. 5 Configuration of the architecture U-Net

MobileNet represented the transfer-learning approach. Its backbone, pretrained on ImageNet, remained frozen during training, while a custom classification head was added, consisting of a global average pooling layer followed by a dense layer with 1024 ReLU units, dropout of 0.5, and a final sigmoid output neuron. Training was carried out for 10 epochs with a batch size of 32 using the Adam optimizer with a learning rate of 1×10^{-4}. This model relied on a stronger geometric augmentation scheme and applied a fixed threshold of 0.5 for binary decisions.

A full fine-tuning strategy was explored with the Vision Transformer (ViT), using the `vit_base_patch16_224` architecture from the `timm` library, pretrained on ImageNet. The entire network was fine-tuned after replacing the classification head with a two-unit linear layer. Training used the Adam optimizer at a learning rate of 1×10^{-4} for 10 epochs with a batch size of 32. No geometric augmentation was applied in order to avoid distortions that could interfere with the model's patch-based representation. Model checkpointing was employed to preserve the weights achieving the highest validation accuracy.

Finally, U-Net was trained from scratch to enable lesion-level segmentation. The model followed a standard encoder–decoder design and was optimized using the Adam optimizer with binary cross-entropy loss. Training was performed for 30 epochs with a batch size of 8, using RGB images as input and producing the corresponding lesion masks as output.

3.5 Statistical analysis

Pairwise model comparison was performed using the McNemar test, appropriate for classifiers that output discrete predictions over the same test samples. Unlike paired t-tests or Wilcoxon tests, McNemar does not assume continuous metric distributions and directly evaluates discordant error patterns between two classifiers.

To estimate metric uncertainty, 95% confidence intervals were computed only for MobileNet and ViT, as they were the top-performing models and the only ones selected for deeper statistical scrutiny. A non-parametric bootstrap with 1000 resamples over the test-set predictions was applied:

$$IC_{95} = [P_{2.5},\ P_{97.5}]$$

Metrics included accuracy, recall, F1-score, and AUC. AlexNet and U-Net were trained only once, and therefore no CI was reported for these architectures.

4 Results

This study evaluated four deep neural network architectures (AlexNet, Vision Transformer, MobileNet, and U-Net) to classify cucumber plant leaves as healthy or diseased. Table 3 summarized the results obtained in terms of accuracy, sensitivity (recall), and F1-score, which are key metrics to evaluate the performance of classification models.

We can observe that the Vision Transformer model demonstrated the best performance in classifying healthy and diseased cucumber leaves, achieving values of 99.39% in accuracy, 99.76% in precision, 99.28% in recall, and 99.52% in F1-score macro. It indicates that the model can correctly classify the leaves, with very few false positives and false negatives. Subsequently, we noted that MobileNet achieved excellent performance with 99% in

Table 3 Model evaluation results

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Loss (%)	AUC (%)
AlexNet	63.00	62.97	64.37	39.00	66.34	50.00
Vision transformer	99.39	99.76	99.28	99.52	4.14	99.43
MobileNet	99.00	99.75	98.26	99.00	3.59	98.92
U-Net	78.23	76.90	61.90	68.59	47.14	84.98

 https://doi.org/10.1007/s00521-026-11945-z

Table 4 Comparison of ViT and MobileNet performance with 95% confidence intervals (CI)

Metric	ViT	MobileNet
Precision	99.53% (98.93–100.00)	98.48% (97.40–99.39)
Recall	99.76% (99.26–100.00)	97.86% (96.43–99.23)
F1-score	99.65% (99.17–100.00)	98.77% (98.04–99.43)
AUC	99.44% (98.75–100.00)	98.70% (97.81–99.41)

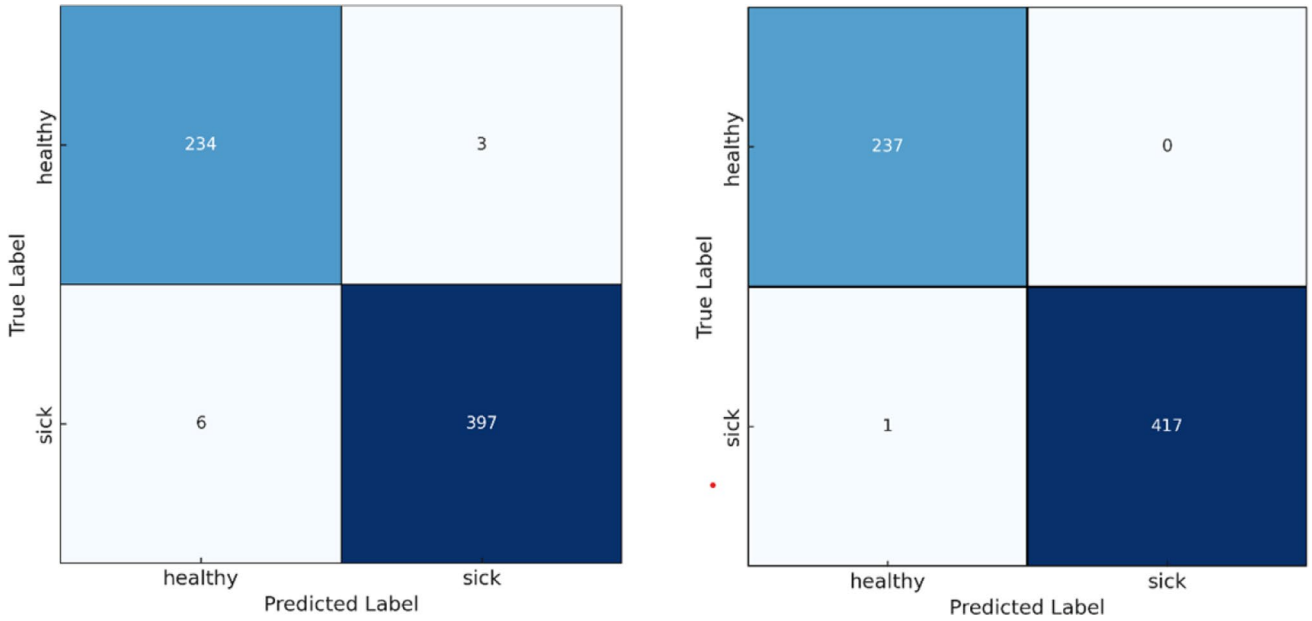


Fig. 6 **a** Confusion matrix of vision transformer from 655 images. **b** Confusion matrix of MobileNet from 655 images.

accuracy, 99.75% precision, 98.26% recall, and 99% F1-score macro. While its performance is slightly inferior to that of the Vision Transformer, it remains highly effective.

In addition to point estimates, 95% confidence intervals were computed for ViT and MobileNet using a non-parametric bootstrap with 1000 resamples, Table 4.

Confidence intervals were not reported for AlexNet and U-Net, since these models were trained and evaluated only once.

On the other hand, U-Net exhibited moderate performance with values of 78.23% accuracy, 76.90% precision, 61.90% recall, and 68.59% F1-score macro. Considering this architecture is primarily helpful for semantic segmentation, it has limitations when classifying complete images. Finally, we observed the lowest performance in classifying healthy and diseased cucumber leaves with the AlexNet architecture, which achieved 63% accuracy, 62.97% precision, 64.37% recall, and 39% F1-score. Its limited capacity to learn complex representations results in more incorrect classifications.

Through the confusion matrix, Fig. 6, we observe that Vision Transformer correctly classified 414 cases as sick when they were indeed sick. Similarly, it accurately identified 237 cases as healthy when they were genuinely healthy. There were no false positives, meaning the model never classified a healthy case as sick. However, there were four false negatives, where the model incorrectly classified sick cases as healthy. The model demonstrates outstanding performance with few errors, exceptionally high sensitivity, and precision in classifying both categories. This architecture has proven to be more robust against variations in the images, such as changes in lighting or capture angles.

From the confusion matrix, Fig. 6, we observe that MobileNet correctly classified 396 cases as sick when they were indeed sick. Additionally, it accurately identified 235 cases as healthy when they were genuinely healthy. There were 2 cases where the model incorrectly classified healthy instances as sick. Similarly, there were 7 cases

where the model incorrectly classified sick instances as healthy. Despite the false negatives, the model still performs well with few errors. However, there is a slight reduction in precision compared to the previous matrix due to an increase in false negative and false positive errors.

Fig. 7a demonstrates that the ViT model has excellent performance on the training set, achieving near 100% accuracy and minimal loss, indicating its strong ability to learn complex patterns in the data. Its transformer-based architecture efficiently captures spatial relationships in images, making it a powerful option for classifying healthy and diseased cucumber leaves. However, the significant difference between the training loss and the validation loss suggests the possibility of overfitting, indicating that the model may be memorizing the data rather than generalizing it correctly. Furthermore, an upward trend in the validation loss is evident in subsequent epochs, thereby underscoring the necessity for optimized regularisation.

The MobileNet model results shown in Fig. 7b demonstrated remarkable proficiency in classifying healthy or diseased cucumber leaves, attaining accuracies that approach 98-100% in both training and validation sets. Despite occasional fluctuations in accuracy across epochs, the similarity between the two curves suggests that the model possesses effective generalization capabilities and exhibits no discernible indications of overfitting. Minor oscillations, training, and validation loss steadily decline, indicating that the model is learning effectively.

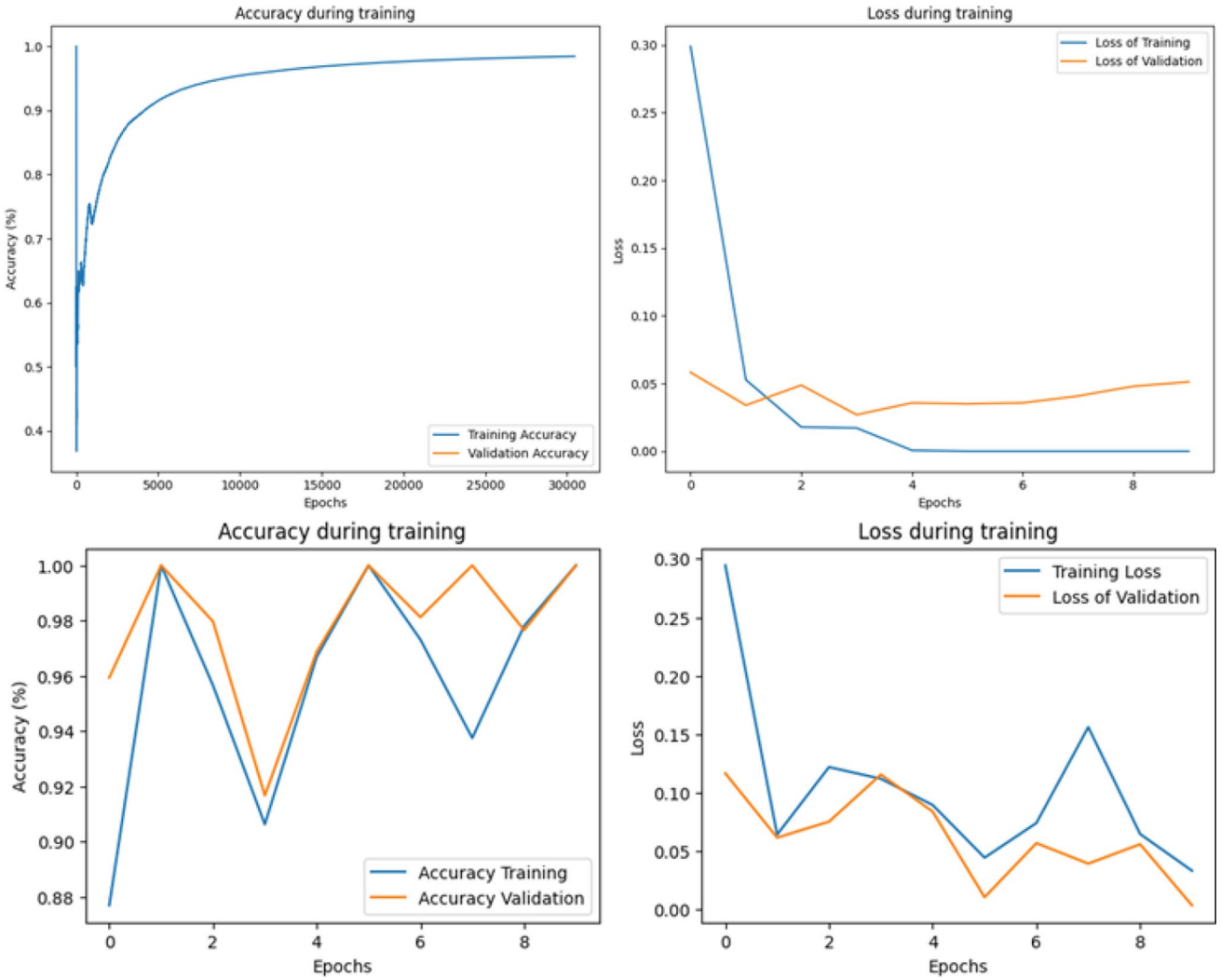


Fig. 7 **a** Learning curve for Vision transformers. **b** Learning curve for MobileNet

Table 5 Contingency table between ViT and MobileNet

	MobileNet	
ViT	Correct	Incorrect
Correct	643	2
Incorrect	9	1

4.1 McNemar test

To assess whether there were significant differences in the classification predictions made by the ViT and MobileNet models, the McNemar test was applied. This statistical test is employed to compare binary classification models applied to the same data set, particularly to analyze the discrepancy between the errors made by each model. Table 5 presents the contingency table generated from the predictions of both models.

- Both models correctly classified item 643 samples.
- 2 samples were correctly classified by MobileNet but not by ViT.
- 9 samples were correctly classified by ViT but not by MobileNet.
- Both models misclassified item 1 sample.

McNemar's test yielded a chi-square statistic of $\chi^2 = 3.2727$ with a p-value of 0.0704. As $p > 0.05$, it confirms the null hypothesis, indicating no statistically significant difference between the error rates of the two models.

5 Discussion

The results obtained allow for interpreting how each architecture responds to the different levels of variability present in the two datasets considered. First, the overall performance indicates that the models adequately capture visual patterns associated with distinguishing between healthy and diseased leaves, but they also exhibit sensitivity to the image acquisition conditions. The contrast between a dataset with a controlled background and another with more realistic scenarios explains the variation observed across architectures: in clean environments, the models prioritize fine textures and local contrast, whereas in field conditions they must learn to ignore structural noise such as shadows, debris, or uneven lighting. This behavior is consistent with recent studies in agricultural vision, where the presence of heterogeneous backgrounds increases the difficulty of extracting discriminative features and can reduce inference accuracy in real-world scenarios [15, 24, 49].

From a methodological standpoint, the differential behavior across architectures can be attributed to their ability to model spatial variability. MobileNet achieves a favorable balance between accuracy and computational cost thanks to its efficient design based on depthwise separable convolutions, which enables better generalization to background noise; this aligns with observations from lightweight agricultural models reported by Wang et al. [41] and Duhan [42]. In contrast, heavier models such as AlexNet or ViT require larger amounts of heterogeneous data to learn robust representations under changes in lighting, pigmentation, or leaf structure an issue also noted in works that evaluate transformers and large CNNs for plant disease detection [11, 19] . This finding aligns with the literature, which reports that lightweight models optimized for mobile devices tend to be more stable when datasets exhibit high unstructured variation, whereas transformers require greater contextual diversity to reach their full potential only when trained on diverse, large-scale datasets [24, 45].

However, although the second dataset introduces a higher degree of realism, both originate from similar environmental contexts, which limits extrapolation to different regions, cultivars, or acquisition devices. Genetic variability in cucumber, nutrient stress, sensor type, and agroclimatic conditions can alter the visual appearance of symptoms, affecting classification consistency. Therefore, real-scale generalization requires incorporating images from multiple farms, phenological stages, and climatic conditions, as well as cross-domain validations that reveal potential biases related to dataset origin.

Regarding practical applicability, the results suggest that MobileNet represents the most viable option for deployment on embedded or low-cost devices, given its reduced number of parameters and low inference time. Although this study did not explicitly evaluate performance on specific hardware, the literature indicates that lightweight models perform better on Raspberry Pi, smartphones, or integrated sensors [41], whereas heavier architectures require additional optimizations. Future stages will incorporate evaluations on representative platforms considering latency, memory consumption, throughput, and energy efficiency, along with compression techniques such as quantization-aware training, pruning, distillation, and export to ONNX/TensorRT.

Finally, the results highlight the need to advance toward strategies that enhance model interpretability and robustness. Likewise, incorporating images under adverse conditions occlusions, multiple co-occurring symptoms, humidity, insects, high reflectance will be essential for strengthening the applicability of the system in real agricultural production scenarios.

6 Conclusion

This work demonstrates the potential of deep neural networks, particularly Vision Transformer and MobileNet, for the early and accurate classification of foliar diseases in cucumber leaves. The results show that these architectures achieve high precision, recall, and F1-score levels and offer significant advantages in terms of computational efficiency and adaptability to environmental variations. Specifically, Vision Transformer stood out for its ability to handle complex data and provide exceptional robustness under field conditions. Its superiority is likely due to its capacity to capture complex spatial relationships and patterns at different scales through attention mechanisms. It is particularly effective in plant leaf images, where textures and visual patterns can indicate diseases.

The integration of advanced preprocessing and data augmentation techniques was essential to optimize the performance of the studied architectures, highlighting the importance of adapting workflows to the specific characteristics of the datasets. Furthermore, the proposed models pave the way for practical applications in innovative agriculture systems, where real-time detection can improve decision-making and minimize economic and environmental losses. The models developed and evaluated in this study are subject to inherent limitations that preclude their generalizability to real agricultural environments. A salient limitation pertains to the potential overfitting to the particular conditions of the original data set, including distinct illumination characteristics, homogeneous backgrounds, or specific leaf types present in the images utilized for training. Another salient limitation pertains to the inherent variability of open-field crops, encompassing factors such as dirt, mechanical damage, insects, the simultaneous occurrence of multiple diseases, and varying plant growth stages. These factors introduce additional complexity we could not fully capture in the datasets.

To address these limitations, future research should consider incorporating images captured under various climatic conditions, varying phenological stages, different soil conditions, and the presence of multiple simultaneous stressors. In addition to evaluating models using data collected from different geographic locations, camera types, drones, or IoT systems to ensure greater generalization to real conditions, we should explore other avenues of research. We also recommend exploring these techniques for multi-label classifications and deploying models on low-cost devices to enhance accessibility in rural agricultural settings. These initiatives would contribute significantly to developing sustainable solutions for global food security.

Acknowledgements The authors express their gratitude to the Call 933 "Training in National Doctorates with a Territorial, Ethnic and Gender Focus in the Framework of the Mission Policy—2023" of the Ministry of Science, Technology and Innovation (Minciencia). In addition, we thank the team of the Artificial Intelligence Laboratory VerbaNex [6], affiliated with the UTB, for their contributions to this project.

Funding Open Access funding provided by Colombia Consortium

[6] https://www.kaggle.com/datasets/kaushigihanml/cucumber-leaf-disease-dataset

Data availability All data supporting the findings of this study are available in the repository [7] and [8].

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Chen X, Shi D, Zhang H, Pérez JAS, Yang X, Li M (2024) Early diagnosis of greenhouse cucumber downy mildew in seedling stage using chlorophyll fluorescence imaging technology. Biosys Eng 242:107–122. https://doi.org/10.1016/J.BIOSYSTEMSENG.2024.04.013
2. Nobel S, Afroj M, Kabir M, Mridha M (2024) Development of a cutting-edge ensemble pipeline for rapid and accurate diagnosis of plant leaf diseases. Artif Intell Agriculture 14:56–72. https://doi.org/10.1016/J.AIIA.2024.10.005
3. Rezaei M, Diepeveen D, Laga H, Jones M, Sohel F (2024) Plant disease recognition in a low data scenario using few-shot learning. Comput Electron Agric 219:108812. https://doi.org/10.1016/J.COMPAG.2024.108812
4. FAO.: How to feed the world in 2050 (2019). https://www.fao.org/fileadmin/templates/wsfs/docs/expert_paper/How_to_Feed_the_World_in_2050.pdf
5. FAO.: FAO: We need to produce more food with less water (2023). https://www.fao.org/panama/noticias/detail/FAO-Necesitamos-producir-m%C3%A1s-alimentos-con-menos-agua/es
6. Completa, R.: Relationship between population growth and food security (2024). https://revistacompleta.com/relacion-entre-crecimiento-poblacional-y-seguridad-alimentaria/
7. Anim-Ayeko AO, Schillaci C, Lipani A (2023) Automatic blight disease detection in potato (*solanum tuberosum* L.) and tomato (solanum lycopersicum, l. 1753) plants using deep learning. Smart Agricultural Tech 4:100178. https://doi.org/10.1016/J.ATECH.2023.100178
8. Anwarul S, Mohan M, Agarwal R (2023) An unprecedented approach for deep learning assisted web application to diagnose plant disease. Procedia Comput Sci 218:1444–1453. https://doi.org/10.1016/J.PROCS.2023.01.123
9. Hadipour-Rokni R, Asli-Ardeh AE, Jahanbakhshi A, Paeen-Afrakoti EI, Sabzi S (2023) Intelligent detection of citrus fruit pests using machine vision system and convolutional neural network through transfer learning technique. Comput Biol Med 155:106611. https://doi.org/10.1016/J.COMPBIOMED.2023.106611
10. Nagachandrika B, Prasath R, Praveen Joe I (2024) An automatic classification framework for identifying type of plant leaf diseases using multi-scale feature fusion-based adaptive deep network. Biomed Sig Proc Control 95:106316. https://doi.org/10.1016/J.BSPC.2024.106316
11. Dai G, Tian Z, Fan J, Sunil CK, Dewi C (2024) Dfn-psan: multi-level deep information feature fusion extraction network for interpretable plant disease classification. Comput Electron Agric 216:108481. https://doi.org/10.1016/J.COMPAG.2023.108481
12. Megalingam R, Gopakumar Menon G, Binoj S, Asandi Sai D, Kunnambath A, Manoharan S (2024) Cowpea leaf disease identification using deep learning. Smart Agricultural Tech 9:100662. https://doi.org/10.1016/J.ATECH.2024.100662
13. Zhao Y, Zhang Z, Wu N, Zhang Z, Xu X (2024) Mafde-dn4: improved few-shot plant disease classification method based on deep nearest neighbor neural network. Comput Electron Agric 226:109373. https://doi.org/10.1016/J.COMPAG.2024.109373
14. Kumar R, Jain A, Sharma V, Midha M, Das P (2024) Enhancing cucumber leaf disease severity prediction through convolutional neural networks and random forest integration. In: 2024 5th Intl Conf Emerging Tech, INCET 2024 . https://doi.org/10.1109/INCET61516.2024.10593479
15. Qiao C, Li K, Zhu X, Jing J, Gao W, Zhang L (2024) Detection of cucumber downy mildew spores based on improved yolov5s. Inform Process Agriculture. https://doi.org/10.1016/J.INPA.2024.05.002

[7] https://www.kaggle.com/code/akshayr009/cucumberdiseases/input

[8] https://github.com/VerbaNexAI

16. Tovar AM, Puertas E, Martinez-Santos JC (2024) Implementation of convolutional neural networks for automated disease detection in cucumber crops. Commn Comput Inform Sci 2208:156–167. https://doi.org/10.1007/978-3-031-75233-912
17. FEN.: Fundacion Española de Alimentacion-CUCUMBER (2018). https://www.fen.org.es/MercadoAlimentosFEN/pdfs/pepino.pdf
18. Agarwal M, Gupta S, Biswas KK (2021) A new conv2d model with modified relu activation function for identification of disease type and severity in cucumber plant. Sustainable Computing Inform Syst 30:100473. https://doi.org/10.1016/J.SUSCOM.2020.100473
19. Cao Y, Chen L, Yuan Y, Sun G (2023) Cucumber disease recognition with small samples using image-text-label-based multi-modal language model. Comput Electron Agric 211:107993. https://doi.org/10.1016/J.COMPAG.2023.107993
20. Shreshtha K, Raj S, Pal A, Tripathi P, Choudhary K, Mitra D, Rani A, Santos-Villalobos S, Tripathi V (2024) Isolation and identification of rhizospheric and endophytic bacteria from cucumber plants irrigated with wastewater: exploring their roles in plant growth promotion and disease suppression. Current Res Microbial Sci 7:100256. https://doi.org/10.1016/J.CRMICR.2024.100256
21. Jin Y, Zhang Y, Lin L, Ying S, Yu C (2024) Cucumber pgip2 is involved in resistance to gray mold disease. Gene 923:148588. https://doi.org/10.1016/J.GENE.2024.148588
22. Muttappagol M, Hiremath S, Vinay Kumar H, Jahir Basha C, Shankarappa K, Venkataravanappa V, Lakshminarayana Reddy C (2024) Multilocus sequence analysis of candidatus phytoplasma asteris associated with phyllody of cucumber in india and development of loop-mediated isothermal amplification (lamp) assay for its detection. Physiol Mol Plant Pathol 133:102350. https://doi.org/10.1016/J.PMPP.2024.102350
23. Sultana N, Shorif S, Akter M, Uddin MS (2023) A dataset for successful recognition of cucumber diseases. Data Brief 49:109320. https://doi.org/10.1016/J.DIB.2023.109320
24. Zhu X, Chen F, Qiao C, Zhang Y, Zhang L, Gao W, Wang Y (2024) Cucumber pathogenic spores' detection using the gcs-yolov8 network with microscopic images in natural scenes. Plant Methods 20(1):1–16. https://doi.org/10.1186/s13007-024-01243-x
25. Li K, Song Y, Zhu X, Zhang L (2024) A severity estimation method for lightweight cucumber leaf disease based on dm-bisenet. Inform Process Agriculture. https://doi.org/10.1016/J.INPA.2024.03.003
26. Liu Y, Wang Z, Wang R, Chen J, Gao H (2023) Flooding-based mobilenet to identify cucumber diseases from leaf images in natural scenes. Comput Electron Agric 213:108166. https://doi.org/10.1016/J.COMPAG.2023.108166
27. Shahapurkar A, Joshi S, Divyashree T, Sasaki K, Jogaiah S (2024) Bioactive carbohydrate from potato peels: a sweet inducer that mediate resistance against powdery mildew disease of cucumber through phytohormonal alterations. Biocatal Agric Biotechnol 57:103102. https://doi.org/10.1016/J.BCAB.2024.103102
28. Hernández I, Gutiérrez S, Tardaguila J (2024) Image analysis with deep learning for early detection of downy mildew in grapevine. Sci Hortic 331:113155. https://doi.org/10.1016/J.SCIENTA.2024.113155
29. Hu Y, Tang J, Yang J (2024) Introducing artificial intelligence technology to plant disease management for sustainable agriculture. Crop Prot 184:106764. https://doi.org/10.1016/J.CROPRO.2024.106764
30. Kalezhi J, Shumba L (2025) Cassava crop disease prediction and localization using object detection. Crop Prot 187:107001. https://doi.org/10.1016/J.CROPRO.2024.107001
31. Mehedi M, Nawer N, Ahmed S, Khan M, Hasib K, Mridha M, Alam M, Nguyen T (2024) Pld-det: plant leaf disease detection in real time using an end-to-end neural network approach based on improved yolov7. Neural Comput Appl 36(34):21885–21898. https://doi.org/10.1007/S00521-024-10409-6
32. Rajpoot V, Tiwari A, Jalal A (2023) Automatic early detection of rice leaf diseases using hybrid deep learning and machine learning methods. Multimed Tools Appl 82(23):36091–36117. https://doi.org/10.1007/S11042-023-14969-Y/FIGURES/12
33. Singh Y, Chaurasia B, Shukla M (2024) Deep transfer learning driven model for mango leaf disease detection. Intl J Syst Assurance Eng Mgmnt 15(10):4779–4805. https://doi.org/10.1007/S13198-024-02480-Y/FIGURES/24
34. Verma S, Chug A, Singh A, Singh D (2023) Pds-mcnet: a hybrid framework using mobilenetv2 with silu6 activation function and capsule networks for disease severity estimation in plants. Neural Comput Appl 35(25):18641–18664. https://doi.org/10.1007/S00521-023-08693-9/TABLES/13
35. El Sakka M, Mothe J, Ivanovici M (2024) Images and cnn applications in smart agriculture. Eur J Remote Sens 57(1):2352386. https://doi.org/10.1080/22797254.2024.2352386
36. Karantoumanis E, Balafas V, Louta M, Ploskas N (2024) Real-time disease detection on bean leaves from a small image dataset using data augmentation and deep learning methods. Soft Comput. https://doi.org/10.1007/s00500-024-10348-3
37. Cheng T, Zhang D, Gu C, Zhou XG, Qiao H, Guo W, Niu Z, Xie J, Yang X (2024) Yolo-cg-hs: a lightweight spore detection method for wheat airborne fungal pathogens. Comput Electron Agric 227:109544. https://doi.org/10.1016/J.COMPAG.2024.109544
38. Zhang M, Yuan C, Liu Q, Liu H, Qiu X, Zhao M (2024) Detection of mulberry leaf diseases in natural environments based on improved yolov8. Forests 15(7):1188. https://doi.org/10.3390/f15071188
39. Bhargava A, Shukla A, Goswami O, Mohammed A, Peerapong U, Monthippa U (2024) Plant leaf disease detection, classification, and diagnosis using computer vision and artificial intelligence: a review. IEEE Access

40. Lin M, Chen M, Zhang Y, Shen C, Ji R, Cao L (2023) Super vision transformer. Int J Comput Vision 131(12):3136–3151. https://doi.org/10.1007/S11263-023-01861-3
41. Wang B, Zhang C, Li Y, Cao C, Huang D, Gong Y (2023) An ultra-lightweight efficient network for image-based plant disease and pest infection detection. Precision Agric 24(5):1836–1861. https://doi.org/10.1007/S11119-023-10020-0
42. Duhan S, Gulia P, Gill NS, Narwal E (2025) Rtr_lite_mobilenetv2: a lightweight and efficient model for plant disease detection and classification. Current Plant Biology 42:100459. https://doi.org/10.1016/J.CPB.2025.100459
43. Rakib AF, Rahman R, Razi AA, Hasan ASMT (2024) A lightweight quantized cnn model for plant disease recognition. Arab J Sci Eng 49:4097–4108. https://doi.org/10.1007/S13369-023-08280-Z
44. Aboelenin S, Elbasheer FA, Eltoukhy MM, El-Hady WM, Hosny KM (2025) A hybrid framework for plant leaf disease detection and classification using convolutional neural networks and vision transformer. Complex Intell Syst 11:142. https://doi.org/10.1007/S40747-024-01764-X/TABLES/6
45. Wei L, Tang J, Chen J, Mukamakuza CP, Zhang D, Zhang T (2025) A lightweight few-shot learning model for crop pest and disease identification. Artif Intell Rev 58(10):329. https://doi.org/10.1007/S10462-025-11323-6
46. Ngugi HN, Akinyelu AA, Ezugwu AE (2024) Machine learning and deep learning for crop disease diagnosis: performance analysis and review. Agronomy 14:3001. https://doi.org/10.3390/AGRONOMY14123001
47. Tariq MU, Saqib SM, Mazhar T, Khan MA, Shahzad T, Hamam H (2025) Edge-enabled smart agriculture framework: integrating iot, lightweight deep learning, and agentic ai for context-aware farming. Results in Eng 28:107342. https://doi.org/10.1016/J.RINENG.2025.107342
48. Sharafat S, Kabya ND, Emu RI, Ahmed MU, Onik JC, Islam MA, Khan R (2025) An iot-enabled ai system for real-time crop prediction using soil and weather data in precision agriculture. Smart Agricultural Tech 12:101263. https://doi.org/10.1016/J.ATECH.2025.101263
49. Kouadio L, Jarroudi ME, Belabess Z, Laasli SE, Roni MZK, Amine IDI, Mokhtari N, Mokrini F, Junk J, Lahlali R (2023) A review on uav-based applications for plant disease detection and monitoring. Remote Sens 15:4273. https://doi.org/10.3390/RS15174273/S1
50. Srinivas S, Sarvadevabhatla R, Mopuri K, Prabhu N, Kruthiventi S, Babu R (2016) A taxonomy of deep convolutional neural nets for computer vision. Frontiers Robot AI 2:172022. https://doi.org/10.3389/FROBT.2015.00036/PDF
51. Chen H, Qin Y, Liu X, Wang H, Zhao J (2024) An improved deeplabv3+ lightweight network for remote-sensing image semantic segmentation. Complex Intell Syst 10(2):2839–2849. https://doi.org/10.1007/S40747-023-01304-Z
52. Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H, Ai G, Brain G Searching for MobileNetV3 (n.d.)
53. Al-hammuri K, Gebali F, Kanan A, Chelvan IT (2023) Vision transformer architecture and applications in digital health: a tutorial and survey. Vis Computing Industry Biomed Art 6(1):1–28. https://doi.org/10.1186/S42492-023-00140-9
54. Peng Y, Sonka M, Chen D (2023) U-Net v2: Rethinking the skip connections of U-Net for medical image segmentation. arXiv: 2311.17791v2
55. Yang W, Xu L, Wang P, Geng D, Li Y, Xu M, Dong Z (2023) More complex encoder is not all you need . arXiv: 2309.11139v3

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Andrea Menco-Tovar[1] · Juan Carlos Martinez-Santos[2] · Edwin Puertas[2]

✉ Andrea Menco-Tovar
amenco@utb.edu.co

Juan Carlos Martinez-Santos
jcmartinezs@utb.edu.co

Edwin Puertas
epuerta@utb.edu.co

[1] School of Engineering, Universidad Tecnológica de Bolívar, Ternera Km 1, Cartagena, Bolívar, Colombia

[2] School of Digital Transformation, Universidad Tecnológica de Bolívar, Ternera Km 1, Cartagena, Bolívar, Colombia