Peertechz

ENGINEERING GROUP

Trends in
Computer Science and Information Technology
OPEN ACCESS JOURNAL

ISSN: 2641-3086 Journal DOI: https://dx.doi.org/10.17352/tcsit

Research Article

Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures

Javanshir Zeynalov[1]*, Yigitcan Cakmak[2], Ishak Pacal[2,3] and Maftun Aliyev[1]

[1]Department of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012, Nakhchivan, Azerbaijan

[2]Department of Computer Engineering, Faculty of Engineering, Igdir University, 76000, Igdir, Turkey

[3]Turkey Department of Computer Engineering, Faculty of Engineering and Architecture, Fenerbahçe University, Istanbul, Turkey

Submitted : 19 March, 2026
Accepted : 30 March, 2026
Published : 31 March, 2026

***Corresponding author**: Javanshir Zeynalov, Department of Electronics and Information Technologies, Faculty of Architecture and Engineering, Nakhchivan State University, AZ 7012, Nakhchivan, Azerbaijan, E-mail: cavansirzeynalov@ndu.edu.az

Keywords: Deep learning; Tomato leaf diseases; Plant disease classification; Transfer learning; Computational efficiency

https://www.engineegroup.us

Check for updates

Abstract

Tomato production is highly vulnerable to foliar diseases that can reduce yield, increase management costs, and complicate timely intervention. Automated image-based diagnosis has therefore become an important research direction for precision agriculture. In this study, we present a comparative evaluation of four representative deep learning backbones for tomato leaf disease classification on the Plant Village dataset: EfficientNetV2-S, ConvNeXt-Base, DeiT3-Base, and Swin-Base. The dataset comprised 18,160 images from ten classes, including nine disease categories and one healthy class, and was divided into training, validation, and test sets using a 70:15:15 split. All models were trained under a standardized transfer learning pipeline with identical preprocessing, augmentation, and optimization settings to enable a fair comparison across architectures. Performance was assessed using accuracy, precision, recall, F1-score, parameter count, and GFLOPs. All evaluated models achieved very high classification performance, with test accuracies of at least 0.9985. Among them, Swin-Base yielded the best overall predictive performance, reaching an accuracy of 0.9989 and an F1-score of 0.9987. In contrast, EfficientNetV2-S provided the most favorable efficiency profile, achieving 0.9985 accuracy with only 20.19 million parameters and 5.4193 GFLOPs. These findings indicate that both convolutional and transformer-based models can deliver highly reliable tomato leaf disease classification under controlled benchmark conditions, while the final model choice should be guided by the application scenario. Swin-Base is preferable when maximum predictive performance is prioritized, whereas EfficientNetV2-S offers a more practical option for computationally constrained deployments.

Introduction

As a vital component of global food security, the tomato (*Solanum lycopersicum*) ranks as one of the most widely cultivated and consumed horticultural crops, with global yields exceeding 186 million tons in 2020 [1,2]. Beyond their immense economic footprint, tomatoes deliver essential vitamins, antioxidants, and nutrients that are crucial to human diets worldwide. Nevertheless, the cultivation of this critical crop faces constant threats from a variety of bacterial, fungal, and viral pathogens, which frequently culminate in substantial financial deficits [3-5]. Viral infections alone, for instance, are responsible for an estimated annual yield reduction of 2% to 10%, while severe outbreaks possess the destructive potential to wipe out entire harvests entirely [6,7].

Historically, identifying these histopathological threats has relied heavily on the naked eye of farmers or agricultural experts [8,9]. This traditional approach, however, is inherently flawed; it is not only labor-intensive but also highly subjective and susceptible to human error, particularly when disease symptoms are faint or closely resemble nutritional disorders and other localized infections [10-12]. Because visual assessments are intrinsically tied to the varying expertise levels of the observer,

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

they frequently lead to incorrect or delayed conclusions. Such diagnostic failures cascade into severe consequences, including the over-application of agrochemicals, escalating financial burdens for growers, and significant ecological harm, thereby highlighting an urgent necessity for rapid, precise, and fully automated disease detection systems [13-15].

Deep learning has emerged as one of the most influential paradigms in artificial intelligence [16], demonstrating remarkable success across a wide range of domains, including medical image analysis, object recognition, scene understanding, industrial inspection, and decision-support systems [17-19]. Its ability to automatically learn discriminative and hierarchical feature representations directly from raw data has substantially reduced the need for handcrafted feature engineering and has enabled major advances in complex pattern recognition tasks [20-24]. Recently, the integration of deep learning methodologies [25,26] into automated plant disease detection has gained considerable momentum, offering a robust and viable solution for disease identification and categorization [27–29]. Within this domain, Convolutional Neural Networks (CNNs) [30,31] have proven exceptionally effective due to their inherent capacity to autonomously derive and process hierarchical features straight from intricate visual data [32,33]. Concurrently, the emergence of Vision Transformers (Vitus), driven by sophisticated self-attention mechanisms, has redefined the benchmarks of state-of-the-art performance across broader computer vision tasks, and these architectures are now being heavily adapted to enhance diagnostic precision within smart agriculture [34,35].

Motivated by these advancements, this study proposes and rigorously evaluates a holistic deep learning framework engineered to automatically classify ten distinct tomato leaf diseases utilizing the widely accessible Plant Village dataset [36]. We conduct a comprehensive performance benchmarking across four cutting-edge neural network architectures: the efficiency-optimized EfficientNetV2-S, the advanced pure transformer DeiT3-Base, the modernized convolutional network ConvNeXt-Base, and the hierarchical vision transformer Swin-Base. The diagnostic efficacy of each model is systematically assessed utilizing standard classification metrics, namely accuracy, precision, recall, and F1-score. Ultimately, by facilitating the creation of accessible, highly accurate, and automated disease management instruments, the insights generated from this research strive to bolster global food security and promote sustainable agricultural practices.

Related work

To provide agricultural workers with an accessible diagnostic utility, Kebir, et al. [37] engineered a highly efficient, bespoke 20-layer Convolutional Neural Network (CNN). By training and validating this model on the Plant Village repository, the researchers successfully classified ten distinct foliage-based tomato pathologies. Ultimately, this streamlined architectural approach delivered a commendable diagnostic accuracy of 97.5%, establishing itself as a highly reliable and practical instrument for targeted disease intervention in farming communities.

Shifting the focus to a previously under-researched pathogen within the deep learning literature, Vásconez, et al. [38] targeted the visual manifestations of bacterial wilt (*Ralstonia solanacearum*). The authors conducted an extensive comparative assessment of fourteen distinct CNN architectures utilizing a proprietary dataset annotated for varying levels of disease severity. Their comprehensive analysis revealed that MobileNet-v2 and Xception yielded the most optimal results, each attaining an accuracy of 97.7% while maintaining an excellent equilibrium between computational speed and diagnostic precision.

To simultaneously capture intricate local details and broader global contexts within foliar images, Chen, et al. [39] devised an innovative hybrid framework that synergizes Convolutional Neural Networks (CNNs) with Transformer-based architectures. Furthermore, the researchers tackled inherent dataset limitations, such as severe class imbalances and data scarcity, by implementing a modified cycle-consistent generative adversarial network (CyTrGAN) to augment their training data. This dual-faceted strategy produced a highly compact yet exceptionally accurate model, registering a 99.45% success rate on the Plant Village dataset.

Similarly exploring architectural fusion, Tiwari, et al. [40] proposed a sophisticated hybrid system integrating a Vision Transformer (ViT) with a traditional Deep Neural Network (DNN) to bolster both the transparency and precision of tomato disease detection. The fundamental innovation in their research lies in an upgraded multi-head self-attention module integrated with an L1-norm attention mechanism, which empowers the network to isolate and prioritize essential morphological traits with greater efficacy. Consequently, this advanced configuration outpaced numerous contemporary methodologies, securing an outstanding accuracy of 99.74% across an extensive experimental dataset.

Addressing the frequent drop in Transformer model efficacy when deployed in uncontrolled, real-world environments, Shehu, et al. [41] formulated three distinct transfer learning paradigms utilizing Vision Transformers (ViTs). To rigorously evaluate the adaptability of their approach across fluctuating environmental parameters, the models were benchmarked against both the standardized Plant Village repository and a novel, field-acquired dataset designated as Tomato Ebola. Empirical outcomes confirmed that their ViT-Base variant not only achieved a 99.17% accuracy on the baseline dataset but also exhibited vastly superior resilience and generalization on the field-collected images when contrasted with alternative deep learning baselines.

Finally, pushing the boundaries of data efficiency, Sun, et al. [42] introduced a pioneering identification framework dubbed EMA-DeiT, which amalgamates a Data-efficient Image Transformer (DeiT) with self-distillation protocols and an exponential moving average (EMA) mechanism. By harnessing these advanced deep learning strategies, the proposed methodology substantially augmented both the stability and the diagnostic accuracy of the network. Ultimately,

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

their architecture yielded exemplary results, realizing a 99.6% accuracy rate on the Plant Village benchmark while concurrently demonstrating remarkable adaptability across an array of heterogeneous datasets.

Materials and methods

Dataset and data preprocessing

This empirical investigation leveraged the widely recognized, open-access PlantVillage repository, a comprehensive database featuring an extensive collection of both pathological and healthy foliage imagery [43]. For the purposes of this study, our analysis was strictly delimited to the tomato leaf subset, which encompasses ten distinct categories: nine specific pathogenic afflictions and a single healthy control group. To visually contextualize the morphological variations and symptomatic expressions inherent to these classifications, a curated selection of representative images is provided in Figure 1. Furthermore, to ensure a robust and unbiased evaluation of the proposed deep learning architectures, the entire image corpus was systematically partitioned into training, validation, and testing subsets utilizing a rigorous 70:15:15 split. A granular breakdown detailing the precise numerical distribution of images across all ten categories and their respective data partitions is systematically cataloged in Table 1.

The preprocessing of raw data constitutes a fundamental prerequisite for engineering robust and highly effective deep learning architectures [44,45]. To guarantee seamless integration with the specific input requirements of the selected pre-trained networks, the entire image corpus was uniformly standardized to a resolution of 224 × 224 pixels. Preceding the computational training phase, pixel intensity values were systematically normalized down to a continuous range of

Table 1: Distribution of the Dataset across Training, Validation, and Test Sets.

Disease/Class Name	Original Total	Train (70%)	Val (15%)	Test (15%)
Bacterial Spot (BS)	2127	1488	319	320
Early Blight (EB)	1000	700	150	150
Healthy (HLT)	1591	1113	238	240
Late Blight (LB)	1909	1336	286	287
Leaf Mold (LM)	952	666	142	144
Tomato Mosaic Virus (TMV)	373	261	55	57
Septoria Leaf Spot (SLS)	1771	1239	265	267
Two-spotted Spider Mite (SM)	1676	1173	251	252
Target Spot (TS)	1404	982	210	212
Tomato Yellow Leaf Curl Virus (TYLCV)	5357	3749	803	805
Total	18160	12707	2719	2734

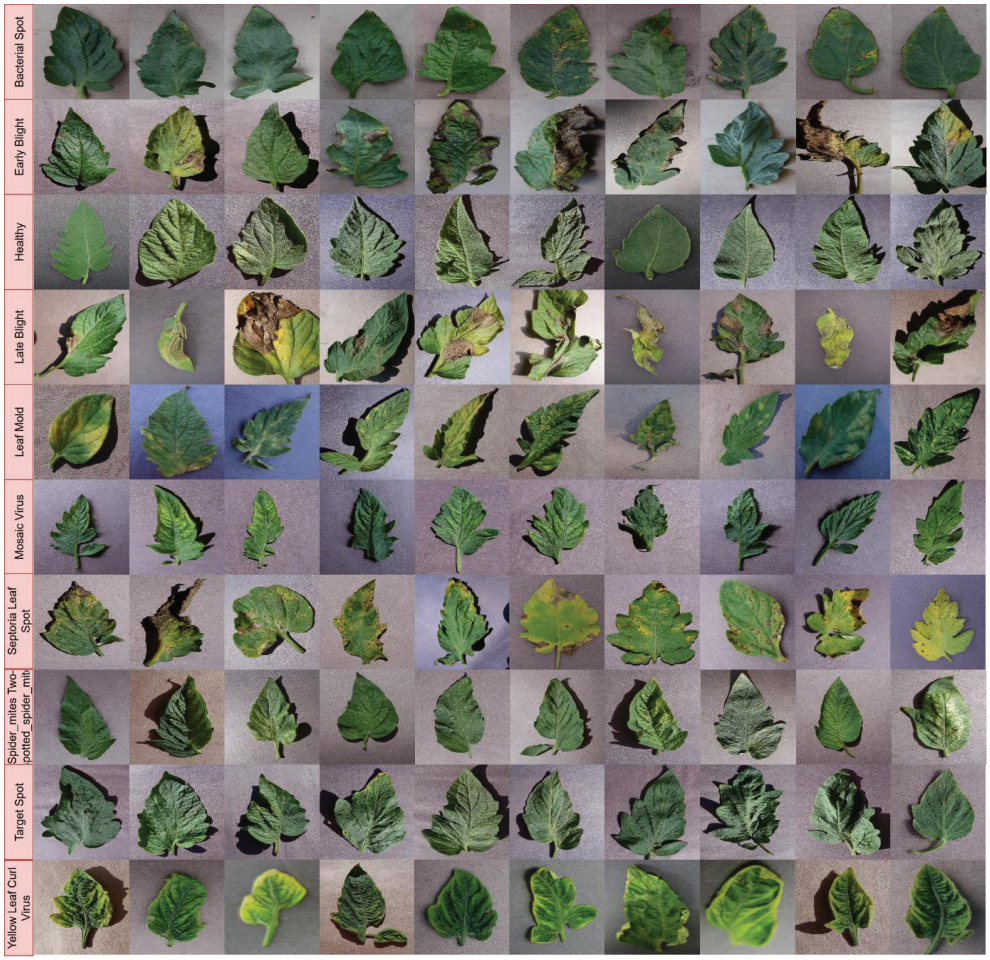

Figure 1: Sample images from the Plant Village dataset illustrating various tomato leaf classes.

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

[0, 1], a critical calibration designed to expedite algorithmic convergence. Moreover, to proactively suppress over fitting tendencies and substantially elevate the models' predictive generalization on unseen data, a rigorous suite of data augmentation strategies was deployed. This augmentation pipeline enriched the training environment by subjecting the original imagery to a variety of stochastic morphological transformations, specifically incorporating random rotational shifts, horizontal inversions, and dynamic zooming operations.

Foundational principles of deep learning models

Convolutional Neural Networks (CNNs): Convolutional Neural Networks (CNNs) have long served as the foundational pillar of modern computer vision, primarily due to their exceptional capacity to autonomously extract and process hierarchical feature representations directly from complex visual data. By employing a series of localized convolutional filters and pooling mechanisms, these networks mimic the human visual cortex, progressively identifying simple geometric edges and textures in their initial layers before synthesizing them into complex, high-level semantic patterns deeper within the architecture. This inherent spatial invariance makes them highly adept at recognizing subtle morphological anomalies in agricultural imagery, such as localized lesions or irregular discoloration on plant foliage. Recent evolutionary steps in this domain have yielded highly optimized, next-generation architectures; for instance, EfficientNetV2-S [46] maximizes diagnostic accuracy while minimizing parameter bloat through progressive learning techniques, whereas ConvNeXt-Base [47] modernizes the standard convolutional framework by integrating advanced structural principles derived from transformers, thereby ensuring robust, state-of-the-art feature extraction while maintaining computational tractability.

Vision Transformers (ViTs): In contrast to the localized, grid-based processing of traditional convolutions, Vision Transformers (ViTs) [48] process images by dividing them into discrete sequences of flattened patches, relying fundamentally on sophisticated self-attention mechanisms to capture complex, global dependencies across the entire visual field. This monumental paradigm shift has redefined the benchmarks of state-of-the-art performance across broader computer vision tasks, and these architectures are now being heavily adapted to enhance diagnostic precision within smart agriculture. By simultaneously weighing the relative importance of all image patches regardless of their spatial distance, ViTs can contextually link disparate disease symptoms across a leaf that might otherwise be overlooked by spatially constrained convolutional filters. Within the scope of this research, we harness advanced transformer variants such as DeiT3-Base [49], a pure transformer model engineered for exceptional data-efficient training, alongside the hierarchical Swin-Base [50], which intelligently reintroduces localized inductive biases through shifted window mechanisms to efficiently and accurately manage high-resolution psychopathological imagery.

Transfer learning and data augmentation strategy

The computational training regimen was executed over a maximum span of 100 epochs utilizing the Adaptive Moment Estimation (Adam) optimizer, strategically chosen for its efficient handling of sparse gradients and adaptive learning rate capabilities. The optimizer was configured with a base learning rate of $1x\ 10^{-4}$ and a weight decay of 2.0×10^{-5} to rigorously penalize complex weights and suppress over fitting. To ensure optimal algorithmic convergence and prevent the models from converging on suboptimal local minima, a cosine annealing learning rate scheduler was integrated into the pipeline. This scheduling included a 5-epoch warm-up phase, initiating at a reduced learning rate of 1.0×10^{-5} alongside a label smoothing factor of 0.1 to meticulously mitigate overconfidence in the network's predictive outputs. Initially, the core computational blocks of the pre-trained convolutional and transformer models were strictly frozen, isolating the back propagation process to train only the newly appended classification layers. Subsequently, these freezing constraints were lifted, allowing the entire network architecture to undergo a meticulous fine-tuning process; this critical phase perfectly aligned the pre-learned, generalized feature extraction mechanisms with the subtle morphological intricacies specific to tomato leaf pathologies.

In synergy with the transfer learning framework, a sophisticated data augmentation pipeline was deployed to artificially amplify the volume and heterogeneity of the training corpus, thereby fortifying the models' diagnostic generalizability against unseen imagery. Moving beyond basic spatial adjustments, the augmentation protocol was meticulously governed by the experimental hyper parameters: stochastic horizontal flipping was applied with a 50% probability, while vertical flipping was deliberately restricted. To simulate the highly variable lighting conditions encountered in real-world agricultural settings, the images were subjected to dynamic color jittering (with a factor of 0.4) to modify brightness, contrast, and saturation. Furthermore, random resized cropping was instituted, dynamically scaling images between 8% and 100% of their original dimensions with fluctuating aspect ratios (ranging from 0.75 to 1.33), supplemented by random spatial interpolations. These computationally generated permutations effectively precluded the network from memorizing superficial spatial orientations or localized background artifacts, compelling the model to strictly assimilate the fundamental, definitive visual biomarkers of the targeted phytopathological conditions.

Experimental design and training protocol

To ensure a rigorous and equitable comparative analysis, the experimental framework was meticulously standardized across all selected deep learning architectures. The entire computational pipeline, encompassing model training, validation, and evaluation, was executed utilizing the Python programming environment integrated with the Tensor Flow backend. To accelerate the computationally intensive training phases and manage complex matrix operations efficiently, all empirical simulations were deployed on a high-performance

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

workstation equipped with a state-of-the-art NVIDIA GeForce RTX 5090 graphic processing unit (GPU).

Maintaining strict methodological consistency, an identical training protocol was uniformly applied to all evaluated networks. Optimization was governed by the Adaptive Moment Estimation (Adam) algorithm, initialized with a learning rate of 1×10^{-4}, to facilitate the comprehensive fine-tuning of the entire architectural pipeline. A batch size of 16 was systematically chosen to establish an optimal equilibrium between memory utilization and algorithmic accuracy over a maximum training duration of 100 epochs. Furthermore, to preemptively counteract over fitting and isolate the most robust iteration of each model, an early stopping mechanism was integrated. This functional safeguard continuously monitored validation loss, terminating the training sequence automatically if no statistical improvement was observed for 10 consecutive epochs. Ultimately, the network weights that yielded the absolute minimum validation loss were preserved and subsequently deployed for the final, unbiased performance appraisal on the isolated testing dataset.

Performance evaluation metrics

To comprehensively quantify both the diagnostic efficacy and the operational viability of the proposed deep learning architectures, a dual-faceted evaluation strategy was implemented. The primary analytical framework relied upon four fundamental statistical indicators, accuracy, precision, recall, and the F1-score, to provide a holistic assessment of the models' predictive performance. In this context, overall accuracy represents the total proportion of correctly classified instances across all tomato leaf categories, whereas precision specifically gauges the system's reliability by measuring the exactness of its positive disease identifications. Concurrently, recall, or sensitivity, ascertains the network's effectiveness in capturing the entirety of actual positive cases. To harmonize the inherent trade-off between precision and recall, a critical necessity for evaluating datasets with potential internal variance, the F1-score is utilized as their harmonic means. The mathematical formulations defining these standard classification criteria are systematically detailed in Equations 1 through 4 below, where the foundational variables represent True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), respectively. Furthermore, because the practical deployment of these diagnostic tools in real-world agricultural settings heavily depends on resource consumption, the models were also strictly evaluated on their computational footprint. Consequently, the architecture was benchmarked against their total number of trainable parameters (Params) and the required Giga Floating-Point Operations per Second (GFLOPs). These specific metrics serve as vital markers of computational efficiency, allowing for a highly nuanced comparison of the inherent trade-offs between a network's structural complexity, its processing cost, and its overall feasibility for deployment.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1} - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Results and discussion

The comprehensive evaluation of the selected deep learning architectures reveals exceptional diagnostic capabilities across all tested models, as systematically detailed in Table 2. The empirical results demonstrate a highly competitive baseline, with EfficientNetV2-S, DeiT3-Base, and ConvNeXt-Base all achieving an identical, outstanding accuracy of 0.9985.

Despite this parity in overall accuracy, subtle variations emerge when examining more granular performance metrics. EfficientNetV2-S recorded a precision of 0.9984, a recall of 0.9983, and an F1-score of 0.9983, showcasing a robust balance in its predictive reliability. In comparison, the pure transformer architecture, DeiT3-Base, yielded a precision of 0.9981, a recall of 0.9972, and an F1-score of 0.9976. Meanwhile, the modernized convolutional network, ConvNeXt-Base, demonstrated slightly superior harmonization with a precision of 0.9984, a recall of 0.9985, and an F1-score of 0.9984. These remarkably high scores across distinct architectural paradigms highlight the profound efficacy of leveraging transfer learning combined with rigorous data augmentation protocols to extract critical Phyto pathological features.

While diagnostic accuracy is paramount, the practical deployment of these models in resource-constrained agricultural environments necessitates a critical evaluation of their computational footprint. As outlined in Table 2, EfficientNetV2-S emerges as the undisputed leader in operational efficiency, requiring a mere 20.19 million parameters and 5.4193 GFLOPs to achieve its near-perfect classification metrics. This extraordinary equilibrium between low computational cost and high predictive power makes it an ideal candidate for integration into mobile or edge-computing diagnostic devices. Conversely, the heavier architectures incur significantly larger computational expenses. DeiT3-Base operates with 85.82 million parameters and 33.6955 GFLOPs, while ConvNeXt-Base utilizes 87.58 million parameters and 30.7075 GFLOPs. Although these advanced models deliver exceptional accuracy, their substantial parameter counts and higher computational processing requirements pose potential integration challenges for real-time, low-power field applications.

Table 2: Model Evaluation Results and Parameter Counts.

Models	Accuracy	Precision	Recall	F1-score	Params	GFLOPs
EfficientNetV2-S	0.9985	0.9984	0.9983	0.9983	20.19	5.4193
DeiT3-Base	0.9985	0.9981	0.9972	0.9976	85.82	33.6955
ConvNeXt-Base	0.9985	0.9984	0.9985	0.9984	87.58	30.7075
Swin-Base	0.9989	0.9988	0.9987	0.9987	86.75	30.3375

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

Amidst the highly competitive models evaluated, the hierarchical Vision Transformer, Swin-Base, distinguished itself by achieving the absolute highest performance metrics across the board. The model registered an unparalleled accuracy of 0.9989, complemented by a precision of 0.9988, a recall of 0.9987, and an F1-score of 0.9987. This superior performance requires 86.75 million trainable parameters and operates at 30.3375 GFLOPs. The exceptional diagnostic precision of the Swin-Base framework can be directly attributed to its unique shifted-window mechanism, which brilliantly synthesizes the global contextual awareness inherent to transformers with the localized feature extraction efficiency typical of convolutional networks. To provide a more granular, class-specific understanding of this optimal model's predictive behavior, its performance on the held-out test set is visually represented in Figure 2.

The confusion matrix corresponding to the Swin-Base architecture's predictions (Figure 2) vividly illustrates the model's formidable capacity to discriminate among complex and visually similar tomato leaf pathologies. An analysis of the matrix reveals virtually flawless classification across the vast majority of the ten categories. Specifically, the model achieved perfect recognition for the Healthy class (240 instances), Leaf Mold (144 instances), Septoria Leaf Spot (267 instances), Two-spotted Spider Mite (252 instances), Tomato Mosaic Virus (57 instances), and Tomato Yellow Leaf Curl Virus (805 instances). The negligible error margin is confined to a mere three misclassifications across the entire test set: one instance of Bacterial Spot (out of 320) was erroneously identified as Late Blight, one Early Blight sample (out of 150) was misclassified as Target Spot, and one Late Blight instance (out of 287) was incorrectly predicted as a Healthy leaf. These exceptionally rare misdiagnoses likely stem from highly ambiguous or overlapping visual symptoms at specific stages of disease progression. Ultimately, these results underscore the remarkable robustness and clinical viability of the Swin-Base architecture for automated agricultural diagnostics.

Conclusion

This study presented a standardized comparative analysis of four contemporary deep learning architectures for tomato leaf disease classification using the PlantVillage dataset. Across all experiments, the evaluated models achieved consistently high performance, confirming that both modern CNN-based and transformer-based backbones are highly effective for this task under controlled imaging conditions. Among the tested models, Swin-Base delivered the strongest predictive results, achieving the highest accuracy and F1-score, while EfficientNetV2-S provided the most attractive balance between classification performance and computational efficiency. These findings suggest that the selection of a model for automated tomato disease diagnosis should not rely on accuracy alone, but should also consider deployment constraints such as parameter count, computational cost, and target hardware.

From a practical perspective, the results support two complementary conclusions. First, hierarchical transformer

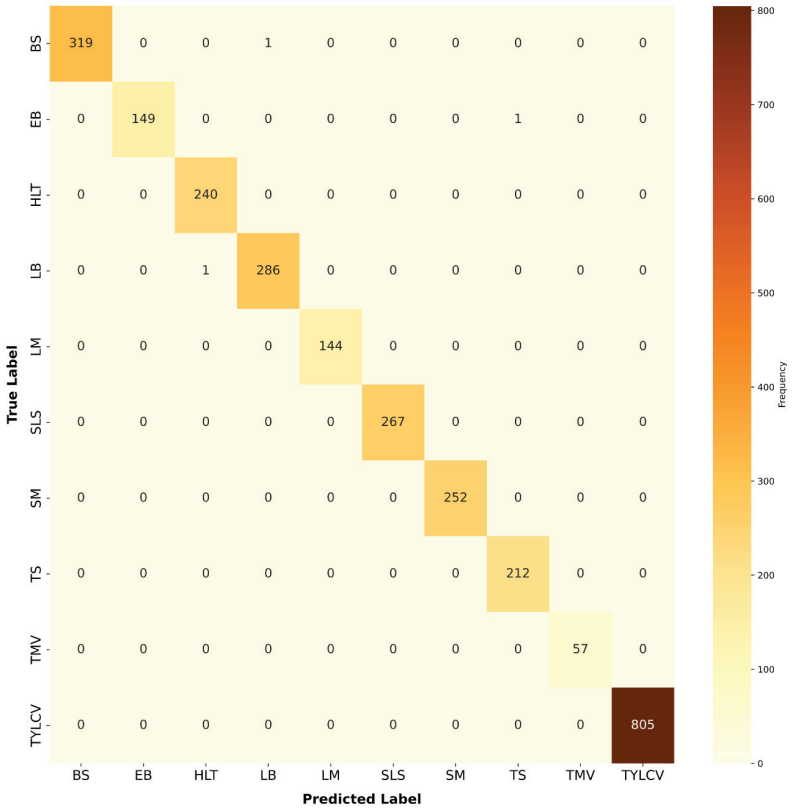

Figure 2: Confusion matrix illustrating the performance of the Swin-Base model on the test set.

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

models such as Swin-Base can provide marginal but meaningful gains in predictive performance when maximum classification accuracy is required. Second, lightweight and efficient CNN-based models such as EfficientNetV2-S remain highly competitive and may be more suitable for edge-oriented or resource-constrained agricultural applications. Accordingly, the present work provides a useful benchmark for selecting appropriate architectures according to different operational needs.

Despite these promising results, the study should be interpreted within the scope of its experimental setting. The models were evaluated on the PlantVillage dataset, which is widely used and suitable for controlled benchmarking, but it does not fully reflect the variability of real field environments, including illumination changes, background complexity, occlusion, and symptom overlap under natural conditions. Future work should therefore focus on external validation using field-acquired datasets, robustness analysis under domain shift, explain ability-based assessment of model decisions, and deployment-oriented testing on mobile or embedded hardware. Such extensions would strengthen the translational value of the proposed benchmarking framework and further support the development of reliable AI-assisted tools for practical disease management in agriculture.

Declarations

Conflict of interest: The authors declare that there are no conflicts of interest regarding the publication of this research.

Declaration of AI and AI-assisted technologies in the writing process: AI-assisted tools were used only for language translation, grammar correction, and improvement of readability during the preparation of this manuscript. No AI tools were used in the study design, methodology, data analysis, model development, result generation, or interpretation of the findings. The authors take full responsibility for the scientific content of the manuscript.

Data availability: The empirical data utilized in this investigation were sourced from the open-access PlantVillage repository.

References

1. Zou F, Hua J, Zhu Y, Deng J, He R. ECVNet: A fusion network of efficient convolutional neural networks and visual transformers for tomato leaf disease identification. Agronomy. 2024;14. Available from: https://doi.org/10.3390/agronomy14122985

2. Chouchane A, Ouamane A, Belabbaci EO, Himeur Y, Amira A. Deep learning-based leaf image analysis for tomato plant disease detection and classification. In: Proceedings - International Conference on Image Processing (ICIP). IEEE Computer Society; 2024. p. 2923–2929. Available from: https://doi.org/10.1109/ICIP51287.2024.10647491

3. Jelali M. Deep learning networks-based tomato disease and pest detection: a first review of research studies using real field datasets. Front Plant Sci. 2024;15:1493322. Available from: https://doi.org/10.3389/fpls.2024.1493322

4. Nemmour H, Mezenner A, Lakehal MR, Bouibed ML, Ammi I, Moghrabi C, et al. GMaxViT: Gabor multi-axis vision transformer for tomato disease identification. 2026:1–6. Available from: https://doi.org/10.1109/aiccsa66935.2025.11315439

5. George R, Nishankar S, Thuseethan S, Wimalasooriya C, Sebastian Y, Ragel RG, Liang Z. U-FedTomAtt: ultra-lightweight federated learning with attention for tomato disease recognition. 2026. Available from: http://arxiv.org/abs/2602.16749 (accessed March 27, 2026).

6. Chandvekar SD, Bhoite S. Machine learning based mobile application for disease identification in tomato leaf. In: Proceedings of the 9th International Conference on Communication and Electronics Systems (ICCES 2024). Institute of Electrical and Electronics Engineers Inc.; 2024. p. 2007–2012. Available from: https://doi.org/10.1109/ICCES63552.2024.10859334

7. Deoli R, Bhatt C, Singh D, Bisht R. Hybrid VGG19–random forest framework for robust tomato leaf disease classification. 2026:1408–1412. Available from: https://doi.org/10.1109/icicnis66685.2025.11315737

8. Hosen MM, Islam MdH. Aggrotech: leveraging deep learning for sustainable tomato disease management. 2025. Available from: https://arxiv.org/abs/2501.12052v1 (accessed October 10, 2025).

9. Pacal I. Enhancing crop productivity and sustainability through disease identification in maize leaves: exploiting a large dataset with an advanced vision transformer model. Expert Syst Appl. 2024;238:122099. Available from: https://doi.org/10.1016/J.ESWA.2023.122099

10. Abdullah A, Amran GA, Tahmid SMA, Alabrah A, AL-Bakhrani AA, Ali A. A deep-learning-based model for the detection of diseased tomato leaves. Agronomy. 2024;14. Available from: https://doi.org/10.3390/agronomy14071593

11. Sunday B, Yusuf SA, Abdullahi M, Masha YA, Matemilola AS. Disorder severity classification in tomato based on weight cluster loss and convolutional neural network. Discover Applied Sciences. 2025;7:1434. Available from: https://doi.org/10.1007/s42452-025-07642-6

12. Ishrak MdA, Mahmud T, Hanip A, Hossain MS. Self-supervised learning for tomato leaf disease classification: a comparative study of SimCLR, MoCo, and BYOL. In: 2025 IEEE 2nd International Conference on Computing, Applications and Systems (COMPAS). 2025. p. 1–6. Available from: https://doi.org/10.1109/COMPAS67506.2025.11381842

13. Kansal S, Jaiswal A, Sachdeva N. Empirical analysis of deep learning models for tomato leaf disease detection. In: Proceedings of the 14th International Conference on Cloud Computing, Data Science and Engineering (Confluence 2024). Institute of Electrical and Electronics Engineers Inc.; 2024. p. 430–435. Available from: https://doi.org/10.1109/Confluence60223.2024.10463386

14. Mo H, Wei L. Tomato yellow leaf curl virus detection based on cross-domain shared attention and enhanced BiFPN. Ecol Inform. 2025;85:102912. Available from: https://doi.org/10.1016/j.ecoinf.2024.102912

15. Sun C, Li Y, Song Z, Liu Q, Si H, Yang Y, et al. Research on tomato disease image recognition method based on DeiT. Eur J Agron. 2025;162:127400. Available from: https://doi.org/10.1016/j.eja.2024.127400

16. Pacal I. Investigating deep learning approaches for cervical cancer diagnosis: a focus on modern image-based models. Eur J Gynaecol Oncol. 2025;46:125–141. Available from: https://www.ejgo.net/articles/10.22514/ejgo.2025.012

17. Bayram B, Ince S, Kilicarslan S, Veziroglu E, Celik O, Pacal I. Advancing ischemic stroke detection through an in-depth evaluation of YOLOv10 models on diffusion-weighted imaging data. J Soft Comput Decis Anal. 2026;4:16–31. Available from: https://doi.org/10.31181/jscda41202680

18. Ganie SM, Pacal I. GBA-NET: a gated bottleneck and attention-driven architecture for robust ischemic stroke segmentation across CT and DWI. Appl Comput Math. 2026;25:52–78. Available from: https://doi.org/10.30546/1683-6154.25.1.2026.52

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106

19. Pacal I. Chaotic learning rate scheduling for improved CNN-based breast cancer ultrasound classification. Chaos Theory Appl. 2025;7:297–306. Available from: https://doi.org/10.51537/CHAOS.1807694

20. Pacal I, Cakmak Y. A comparative analysis of U-Net-based architectures for robust segmentation of bladder cancer lesions in magnetic resonance imaging. Eurasian J Med Oncol. 2025;9(4):268–283. Available from: https://doi.org/10.36922/EJMO025260276

21. Suhani S, Gopi AP, Naik KJ. A quantization training based deep learning for plant disease detections, severity and yield-loss prediction in sustainable agriculture. Soft Comput. 2026;30:2677–2699. Available from: https://doi.org/10.1007/s00500-025-11150-5

22. Cakmak Y, Pacal I. Comparative analysis of transformer architectures for brain tumor classification. Open Exploration. 2025;6:1001377. Available from: https://doi.org/10.37349/emed.2025.1001377

23. Naral S, Cakmak Y, Pacal I. Comparative evaluation of vision transformers and convolutional networks for breast ultrasound image classification. Open Exploration. 2026;7:1001382. Available from: https://doi.org/10.37349/emed.2026.1001382

24. Devarajan D, Allafi R, Obayya M, Nemri N. AI based real time disease diagnosis in plants using deep learning driven CNNs. Sci Rep. 2026;16:4587. Available from: https://doi.org/10.1038/s41598-025-34681-1

25. Cakmak Y, Zeynalov J. A comparative analysis of convolutional neural network architectures for breast cancer classification from mammograms. Artif Intell Appl Sci. 2025;1:28–34. Available from: https://doi.org/10.69882/ADBA.AI.2025075

26. Cakmak Y. Machine learning approaches for enhanced diagnosis of hematological disorders. Comput Syst Artif Intell. 2025;1:8–14. Available from: https://doi.org/10.69882/ADBA.CSAI.2025072

27. Sun Y, Ning L, Zhao B, Yan J. Tomato leaf disease classification by combining EfficientNetv2 and a Swin transformer. Appl Sci (Switzerland). 2024;14. Available from: https://doi.org/10.3390/app14177472

28. Praneesh M, Kavitha V, Santhakumari R, Shobana R. Exploring incremental and deep learning approaches for plant disease detection: a comprehensive review. Real-Time Artificial Intelligence (AI). 2025:31–54. Available from: https://doi.org/10.1201/9781998511358-2

29. Khubisa FM, Olugbara OO. Bibliometric analysis of deep learning in plant disease management. Egypt Inform J. 2026;33:100880. Available from: https://doi.org/10.1016/j.eij.2025.100880

30. Cakmak Y, Pacal N. Deep learning for automated breast cancer detection in ultrasound: a comparative study of four CNN architectures. Artif Intell Appl Sci. 2025;1:13–19. Available from: https://doi.org/10.69882/ADBA.AI.2025073

31. Cakmak Y, Maman A. Deep learning for early diagnosis of lung cancer. Comput Syst Artif Intell. 2025;1:20–25. Available from: https://doi.org/10.69882/ADBA.CSAI.2025074

32. Ahmed MR, Haque R, Rahman SMA, Afridi S, Abir MFF, Hossain MF, et al. Towards automated detection of tomato leaf diseases. In: Proceedings - 6th International Conference on Electrical Engineering and Information and Communication Technology (ICEEICT 2024). Institute of Electrical and Electronics Engineers Inc.; 2024. p. 387–392. Available from: https://doi.org/10.1109/ICEEICT62016.2024.10534559

33. Zeynalov J, Cakmak Y, Pacal I. Automated apple leaf disease classification using deep convolutional neural networks: a comparative study on the Plant Village dataset. J Comput Sci Digit Technol. 2025;1:5–17. Available from: https://doi.org/10.5281/zenodo.17204089

34. Poly A, Vetriveeran D, Balamurugan M. Tomato plant disease classification using transfer learning. In: 2024 5th International Conference for Emerging Technology (INCET 2024). Institute of Electrical and Electronics Engineers Inc.; 2024. Available from: https://doi.org/10.1109/INCET61516.2024.10593634

35. Hoang TM, Bui VH, Nguyen VS, Doan DT, Dang HA, Pham AT. A comprehensive evaluation of lightweight deep learning models for tomato disease classification on edge computing environments. Sci Rep. 2026;2026. Available from: https://doi.org/10.1038/s41598-026-42439-6

36. Bhosale S, Chhabria A. Tomato plant disease identification via deep learning technique. 2024;26. Available from: https://doi.org/10.1142/S0219467826500075

37. Kebir ST, Berrhail F, Didi F. An efficient tomato diseases detection and classification methodology using CNN deep learning network. Braz J Technol. 2024;7:e69316. Available from: https://doi.org/10.38152/bjtv7n2-002

38. Vásconez JP, Vásconez IN, Moya V, Calderón-Díaz MJ, Valenzuela M, Besoain X, et al. Deep learning-based classification of visual symptoms of bacterial wilt disease caused by Ralstonia solanacearum in tomato plants. Comput Electron Agric. 2024;227. Available from: https://doi.org/10.1016/j.compag.2024.109617

39. Chen Z, Wang G, Lv T, Zhang X. Using a hybrid convolutional neural network with a transformer model for tomato leaf disease detection. Agronomy. 2024;14. Available from: https://doi.org/10.3390/agronomy14040673

40. Tiwari M, Kumar H, Prakash N, Kumar S, Neware R, Tripathi S, et al. Tomato disease detection using vision transformer with residual L1-norm attention and deep neural networks. Int J Intell Eng Syst. 2024;17:679–688. Available from: https://doi.org/10.22266/ijies2024.0229.57

41. Shehu HA, Ackley A, Marvellous M, Eteng OE. Early detection of tomato leaf diseases using transformers and transfer learning. Eur J Agron. 2025;168. Available from: https://doi.org/10.1016/j.eja.2025.127625

42. Sun C, Li Y, Song Z, Liu Q, Si H, Yang Y, et al. Research on tomato disease image recognition method based on DeiT. Eur J Agron. 2025;162. Available from: https://doi.org/10.1016/j.eja.2024.127400

43. PlantVillage dataset. Available from: https://www.kaggle.com/datasets/emmarex/plantdisease (accessed October 10, 2025).

44. Wang Z, Wang P, Liu K, Wang P, Fu Y, Lu CT, et al. A comprehensive survey on data augmentation. 2024. Available from: https://arxiv.org/abs/2405.09591v3 (accessed May 28, 2025).

45. Mumuni A, Mumuni F, Gerrar NK. A survey of synthetic data augmentation methods in machine vision. Mach Intell Res. 2024;21:831–869. Available from: https://doi.org/10.1007/S11633-022-1411-7

46. Tan M, Le QV. EfficientNetV2: smaller models and faster training. 2021. Available from: https://arxiv.org/abs/2104.00298 (accessed September 25, 2025).

47. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. Available from: https://github.com/facebookresearch/ConvNeXt (accessed July 23, 2025).

48. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. (accessed August 20, 2025). Available from: https://doi.org/10.48550/arXiv.2010.11929

49. Touvron H, Cord M, Jégou H. DeiT III: revenge of the ViT. 2022;17:13. Available from: https://doi.org/10.48550/arXiv.2204.07118

50. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: hierarchical vision transformer using shifted windows. (accessed July 23, 2025). Available from: https://doi.org/10.48550/arXiv.2103.14030

Citation: Zeynalov J, Cakmak Y, Pacal I, Aliyev M. Deep Learning for Tomato Leaf Disease Classification: Comparative Benchmarking of CNN and Vision Transformer Architectures. Trends Comput Sci Inf Technol. 2026;11(1):027-034. Available from: https://dx.doi.org/10.17352/tcsit.000106