# Precision Agriculture: Tomato Disease Classification via Compact Convolutional Vision Transformer

Amine Mezenner(1), Mohamed Rayane Lakehal(1), Naouel Arab(1), Hassiba Nemmour(1)*, Youcef, Chibani(1)
(1) Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants (LISIC),
Faculty of Electrical Engineering, University of Sciences and Technology Houari Boumediene (USTHB), Algiers, Algeria
**nemmour.hassiba@gmail.com

**Abstract:** Plant disease detection is a one of the most studied subjects in precision agriculture which aims to protect and improve agricultural crops. Commonly, intelligent systems based on CNN (Convolutional Neural Networks) are employed to identify multiple plant diseases by analyzing leaf images. In this work, we propose the use of the Compact Convolutional vision Transformer for tomato disease classification. Experiments conducted on a set of 10 tomato disease categories highlight the effectiveness of the proposed system which outperforms famous CNN models including DenseNet201, and MobileNetV2 by 1.73% in the overall classification accuracy.
.

**Keywords** Convolutional Neural Networks, compact convolutional transformer, tomato disease classification.

## 1. INTRODUCTION

Early detection of plant leaf diseases is crucial for improving crop production and preserving the environment by minimizing the excessive and often indiscriminate use of chemical pesticides. Such practices, when uncontrolled, not only burden the agricultural ecosystem but also pose significant risks to human health. Traditionally, diagnosing plant diseases was the domain of skilled human experts, requiring manual examination of visual symptoms such as leaf discoloration, lesions, or other anomalies. While effective to an extent, this approach is labor-intensive, time-consuming, and highly prone to human error, especially when dealing with large-scale agricultural operations. These challenges have spurred researchers to explore intelligent systems leveraging machine learning and image processing techniques to perform automatic diagnoses of plant leaf diseases with greater speed and accuracy.

Typically, these intelligent systems consist of three main stages: preprocessing, feature extraction, and classification [1]. The preprocessing stage is critical for enhancing image quality and preparing the data for subsequent analysis. It encompasses a variety of enhancement techniques, including segmentation to isolate the leaf from the background, filtering to reduce noise, and background elimination to focus solely on the relevant features of the leaf. Effective preprocessing ensures that the subsequent feature extraction and classification processes operate on high-quality input data, thereby improving the overall system's reliability.

Feature extraction, the second stage, involves deriving meaningful attributes from the processed image to represent the health status of the leaf [2]. These features can be broadly categorized into three types: color, texture, and gradient features. Color features, often represented as histograms from RGB or HSB (Hue, Saturation, and Brightness) color spaces, provide valuable insights into changes in pigmentation that may indicate disease. Texture features, on the other hand, capture patterns and spatial arrangements within the leaf surface using techniques like the gray-level co-occurrence matrix (GLCM) and local binary patterns (LBP). These methods quantify variations in leaf surface roughness and regularity, which are indicative of disease progression. Gradient features, such as those obtained through the histogram of oriented gradients (HOG) and the scale-invariant feature transform (SIFT), provide information about edge orientations and geometric structures within the image,

further enhancing the robustness of the feature representation [3, 4].
In the final classification stage, machine learning algorithms are employed to assign labels to the extracted features, indicating the presence or absence of a particular disease. Early approaches utilized traditional methods like artificial neural networks (ANN) and Support Vector Machines (SVM). However, these methods often relied on handcrafted features, which limited their ability to generalize across diverse datasets. The advent of deep learning revolutionized this stage, with Convolutional Neural Networks (CNNs) emerging as the de facto standard. CNNs possess the unique capability to integrate feature extraction and classification into a unified framework, significantly enhancing their performance on complex image-based tasks [5, 6].They have been widely applied across diverse domains, such as steel surface defect detection and COVID-19 image and video denoising, further demonstrating their effectiveness in image classification and detection tasks [7,8]. Several well-known CNN architecture such as VGG [9], ResNet [10], DenseNet [11], and MobileNet [12] have been successfully adopted in various applications, including plant leaf classification.
Recently, Vision Transformers (ViT) have gained attention as robust classification systems capable of outperforming CNNs in certain scenarios [13, 14]. Unlike CNNs, which rely on convolutional operations, ViTs leverage a transformer-based architecture that excels at capturing global dependencies within the image. Basic ViT models consist of several key components, including patch embedding layers, linear normalization, multi-head attention mechanisms, and fully connected layers. These components work together to process the image as a sequence of patches, enabling the model to learn intricate relationships between different regions of the image. For plant leaf disease classification, several improved ViT implementations have been explored. For instance, in [13], the authors conducted a comparative analysis of basic ViT and Swin Transformer models, achieving an overall classification accuracy of 95.22% on the PlantVillage tomato dataset using the basic ViT. This level of accuracy is comparable to state-of-the-art CNN models and classical systems based on handcrafted features and SVM [4].
Building on these advancements, the present work introduces a novel deep learning ensemble framework for tomato disease classification. Our approach combines the strengths of multiple deep learning models, specifically leveraging deep features derived from Compact Convolutional Transformers (CCT) and CNN models. By fusing the complementary feature representations from these architectures, our ensemble aims to achieve superior classification performance, addressing the limitations of individual models and paving the way for more reliable and efficient plant disease detection systems.

## 2. PROPOSED SYSTEM

Presently, we propose the use of the Compact Convolutional Transformer (CCT) for tomato disease classification. The pipeline of this system is illustrated in Fig.1. The CCT represents an evolution of the Vision Transformer (ViT) architecture, designed to address certain limitations of its predecessor, particularly in handling smaller datasets or images with less pronounced features. The ViT is a robust classification system composed of several processing blocks, including positional embedding, image tokenization, linear projections, transformer encoder, multi-head attention mechanisms, and a fully connected feed-forward network for the final classification step [1 ,3]. While powerful, the classical ViT architecture often requires large amounts of data and computational resources to achieve optimal performance.

To overcome these challenges, several improved architectures have been proposed, with the Compact Convolutional Transformer (CCT) standing out as a notable enhancement. The CCT introduces a convolutional tokenization process that replaces the traditional image patching technique used in ViT. Instead of dividing the image into fixed patches, convolutional tokenization generates tokens with richer information by preserving spatial relationships and capturing local features more effectively. This approach allows the model to extract finer-grained details, which is particularly advantageous for plant disease classification, where subtle visual cues can be critical for accurate diagnosis.
Additionally, the CCT replaces the classical class token mechanism in the ViT with a sequence pooling process [15, 16]. This adjustment eliminates the dependency on a single class token to represent the entire image, instead leveraging a pooling operation across all tokens to generate a more robust

representation of the input image. By integrating convolutional layers and sequence pooling, the CCT combines the strengths of convolutional networks and transformer-based architectures, enabling it to operate effectively on smaller datasets while maintaining high classification accuracy.

In our proposed approach, the CCT architecture is tailored to address the specific challenges associated with tomato disease classification. Tomato plant leaves exhibit a wide range of visual patterns depending on the type and severity of the disease, necessitating a model that can capture both global and local features with precision. The convolutional tokenization in the CCT ensures the extraction of rich local details, while the transformer encoder leverages multi-head attention to learn complex interrelationships between tokens. This combination makes the CCT particularly well-suited for analyzing the PlantVillage tomato dataset, which includes diverse samples across multiple disease categories.

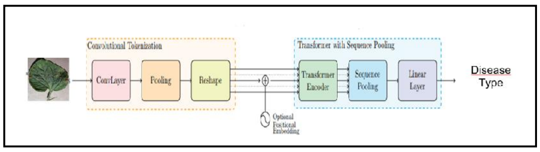


Fig. 1 Pipeline of the compact convolutional transformer for plant leaf disease classification

## 3. EXPERIMENTAL RESULTS

Experiments were conducted on the PlantVillage tomato dataset, which consists of images representing both healthy tomato leaves and nine common tomato pathologies. These pathologies include Target Spot, Mosaic Virus, Yellow Leaf Curl Virus, Bacterial Spot, Early Blight, Late Blight, Leaf Mold, Septoria Leaf Spot, and Spider Mites. The dataset contains a total of 16,011 samples, with the number of images per class ranging from 373 to 3,208. To ensure robust model evaluation, the dataset was split into two subsets: 80% of the samples were used for training, while the remaining 20% were allocated for validation. Figure 2 illustrates representative samples from the dataset, showcasing the diversity of disease patterns and healthy leaf appearances

.

Fig. 2 Samples from experimental dataset (From left to right: Target Spot, Mosaic Virus, Yellow Leaf Curl Virus, Bacterial Spot, Early blight, Healthy, Late blight, Leaf Mold, Septoria Leaf Spot, Spider Mites).

To assess the effectiveness of the proposed system, we compared the performance of the Compact Convolutional Transformer (CCT) with two state-of-the-art convolutional neural networks, DenseNet201 and MobileNetV2. The results are summarized in Table 1.

The CCT demonstrated superior classification performance across several disease categories. Specifically, it achieved significant improvements in classes such as Early Blight, Late Blight, Leaf Mold, Target Spot, and Healthy leaves. Notably, the CCT achieved a total accuracy of 98.37%, outperforming DenseNet201 by 1.73% and MobileNetV2 by 2.43%. These results highlight the CCT's ability to capture both local and global features effectively, making it particularly well-suited for this complex classification task.

Table 1 Results obtained for tomato disease classification

| **Class** | **Accuracy** | | |
|---|---|---|---|
| | **DenseNet201** | **MobileNetV2** | **CCT** |
| **Bacterial Spot** | 99.06 | 98.59 | 96.00 |
| **Early Blight** | 85.50 | 84.00 | 93.00 |
| **Late Blight** | 96.34 | 94.24 | 99.48 |
| **Leaf Mold** | 89.47 | 95.79 | 98.95 |
| **Septoria Leaf Spot** | 99.44 | 94.35 | 98.87 |
| **Two Spotted Spider Mite** | 96.70 | 94.29 | 97.30 |
| **Target Spot** | 93.24 | 93.59 | 98.93 |
| **Yellow Leaf Curl Virus** | 99.69 | 100.0 | 99.84 |
| **Mosaic Virus** | 98.67 | 97.33 | 100.0 |
| **Healthy** | 96.54 | 98.74 | 100.0 |
| **Total Accuracy** | 96.64 | 95.94 | 98.37 |

## 4. CONCLUSION

In this work, we proposed the use of the CCT (Compact Convolutional Transformer) for tomato leaf disease classification. This system was evaluated in comparison with two famous CNN models that are MobileNetV2, and the DenseNet201.The results demonstrated the superiority of the CCT over the CNN models, particularly in seven of the ten disease categories. The CCT achieved a total accuracy of 98.37%, yielding an improvement of 1.73% over DenseNet201 and 2.43% over MobileNetV2. Notable performance gains were observed in challenging classes such as Early Blight, Leaf Mold, and Target Spot, where the CCT effectively captured subtle disease features. These findings highlight the effectiveness of the CCT as a robust and efficient model for tomato leaf disease classification. Its ability to outperform traditional CNN architectures makes it a promising candidate for real-world applications in precision agriculture, where accurate and timely disease diagnosis is critical for optimizing crop health and productivity.

### *References*


[1] A. Mezenner, H. Nemmour, Y. Chibani, and A. Hafiane, “Tomato plant leaf disease classification based on CNN features and support vector machines,” in Proc. Int. Conf. Adv. Electr. Eng., 2022, pp. 1–5.

[2] J. Basavaiah and A. Arlene Anthony, “Tomato leaf disease classification using multiple feature extraction techniques,” Wireless Pers. Commun., vol. 115, no. 1, pp. 633–651, 2020.

[3] A. Mezenner, H. Nemmour, Y. Chibani, and A. Hafiane, “Local directional patterns for plant leaf disease detection,” in Proc. Int. Conf. Adv. Electron., Control Commun. Syst., 2023, pp. 1–5.

[4] T. Hayıt, H. Erbay, F. Varcın, F. Hayıt, and N. Akci, “The classification of wheat yellow rust disease based on a combination of textural and deep features,” Multimedia Tools Appl., vol. 82, pp. 47405–47442, 2023.

[5] A. Mezenner, H. Nemmour, Y. Chibani, and A. Hafiane, “Comparison of CNN-based features with gradient features for tomato plant leaf disease detection,” in Multimedia Data Processing and Computing, CRC Press: Taylor & Francis Group, 2023, pp. 32–42, ISBN: 9781003391272.

[6] M. Agarwal, S. K. Gupta, and K. K. Biswas, “Development of efficient CNN model for tomato crop disease identification,” Sustain. Comput.: Informatics Syst., vol. 28, Art. no. 100407, 2020.

[7] Y. Kateb, H. Meglouli, and A. Khebli, “Steel surface defect detection using convolutional neural network,” Algerian J. Signals Syst., vol. 5, no. 4, pp. 203–208, 2020.

[8] Z. Messali, S. S. Saoud, and A. Lamreche, “COVID-19 images and video denoising algorithms based on convolutional neural networks (CNNs),” Algerian J. Signals Syst., vol. 6, no. 2, pp. 122–129, 2021.

[9] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” arXiv preprint, arXiv:1312.6034, 2013.

[10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.

[11] G. Huang, S. Liu, L. Van der Maaten, and K. Q. Weinberger, “CondenseNet: An efficient DenseNet using learned group convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 2752–2761.

[12] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” arXiv preprint, arXiv:1704.04861, 2017.

[13] W. Moonwar et al., “Tomato leaf disease classification with vision transformer variants,” in Pattern Recognition, H. Lu, M. Blumenstein, S. B. Cho, C. L. Liu, Y. Yagi, and T. Kamiya, Eds. Cham, Switzerland: Springer, 2023, Lect. Notes Comput. Sci., vol. 14406.

[14] H. Nemmour, A. Mezenner, N. Arab, M. R. Lakehal, Z. Khelfa, and Y. Chibani, “Deep learning ensemble based on transformer and CNN features for tomato leaf disease classification,” in Proc. 5th Symp. Pattern Recognit. Appl. (SPRA 2024), vol. 13540, SPIE, Feb. 2025, pp. 61–66.

[15] A. Hassani, S. Walton, N. Shah, A. Abuduweili, J. Li, and H. Shi, “Escaping the big data paradigm with compact transformers,” Trans. Mach. Learn. Res., pp. 1–22, 2022.

[16] A. I. Jajja, A. Abbas, H. A. Khattak, G. Niedbała, A. Khalid, H. T. Rauf, and S. Kujawa, “Compact convolutional transformer (CCT)-based approach for whitefly attack detection in cotton crops,” Agriculture, vol. 12, no. 10, Art. no. 1529, 2022.