International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 2, April 2025, pp. 2334~2344
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i2.pp2334-2344

Plant disease detection using vision transformers

Mhaned Ali, Mouatassim Salma, Mounia El Haji, Benhra Jamal
Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE), Team OSIL, Department GIL, National High School of Electrical and Mechanical Engineering (ENSEM), University Hassan II, Casablanca, Morocco

Article Info

Article history:

Received Jun 21, 2024
Revised Nov 23, 2024
Accepted Dec 2, 2024

Keywords:

Automated disease classification
Deep learning in agriculture
Image classification
Plant disease detection
Precision agriculture
Vision transformers

ABSTRACT

Plant diseases present a major risk to worldwide food security and the sustainability of agriculture, leading to substantial economic losses and hindering rural livelihoods. Conventional methods for disease detection, including visual inspection and laboratory-based techniques, are limited in their scalability, efficiency, and accuracy. This paper addresses the critical problem of accurately detecting and diagnosing plant diseases using advanced machine learning techniques, specifically vision transformers (ViTs), to overcome these limitations. ViTs leverage self-attention mechanisms to capture intricate patterns in plant images, enabling accurate and efficient disease classification. This paper reviews the literature on deep learning techniques in agriculture, emphasizing the growing interest in ViTs for plant disease detection. Additionally, it presents a comprehensive methodology for training and evaluating ViT models for plant disease classification tasks. Experimental results demonstrate the effectiveness of ViTs in accurately identifying various plant diseases across a balanced 55 classes dataset, highlighting their potential to revolutionize precision agriculture and promote sustainable farming practices.

Corresponding Author:

Mhaned Ali
Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE), Team OSIL, Department GIL, National High School of Electrical and Mechanical Engineering (ENSEM), University Hassan II
Casablanca, Morocco
Email: ali.mhaned.doc21@ensem.ac.ma

1. INTRODUCTION

Agriculture plays a fundamental role in sustaining human life by providing food, feed, fiber, and fuel essential for survival [1]. Sustainable agricultural practices are crucial for ensuring food security, enhancing environmental quality, and maintaining the economic viability of farms [2]. The importance of agriculture in improving the social and economic well-being of individuals and communities has been widely recognized [3]. By adopting sustainable practices, agriculture can contribute to the conservation of ecosystems, protect soil health, and ensure the availability of resources for future generations. Sustainable agriculture also involves balancing economic, social, and environmental aspects to support decision-making and long-term agricultural productivity.

Plant disease detection in agriculture is crucial for maintaining crop yield, food security, and economic sustainability. Utilizing advanced technologies such as machine learning, deep learning, and computer vision has been emphasized in various studies for efficient disease detection in plants [4]–[6]. Traditional methods for plant disease detection have historically relied on visual inspection by experts and laboratory-based techniques. Visual inspection involves examining plants for visible symptoms of diseases, such as lesions, discoloration, or deformities. This method, while commonly used, has limitations in terms of scalability, efficiency, and accuracy

[7]. Visual inspection is subjective and dependent on the expertise of the inspector, leading to variability in results [8]. Additionally, visual inspection may only detect diseases once symptoms are visible, potentially missing asymptomatic infections [9]. These methods offer higher specificity and can detect diseases even in the absence of visible symptoms. However, laboratory-based techniques are time-consuming, expensive, and require specialized equipment and trained personnel, making them less scalable and efficient for large-scale disease surveillance [7]. The limitations of visual inspection and laboratory-based techniques underscore the necessity for more advanced and automated approaches for plant disease detection. Technologies like high-throughput sequencing and artificial intelligence have demonstrated potential in improving the speed, accuracy, and scalability of disease detection [10], [11]. By integrating these advanced technologies with traditional methods, it is feasible to overcome the constraints of visual inspection and laboratory-based techniques, leading to more effective plant disease management strategies.

Internet of things (IoT) and artificial intelligence (AI) technologies are significantly enhancing agricultural practices, particularly in irrigation management. In Morocco, where agriculture heavily relies on rainfall, traditional irrigation methods often lead to water wastage and suboptimal crop hydration. To address this, intelligent irrigation systems using Node-MCU 32S boards monitor air temperature, humidity, soil moisture, and light. This data is sent via MQTT to a Raspberry Pi, where long short-term memory (LSTM) neural networks analyze historical weather data to forecast crop water needs and determine the precise irrigation requirements [12]. Another system uses wireless sensor networks (WSN) and IoT to automate irrigation, minimizing human intervention and water consumption. Data from soil moisture and weather sensors is sent to ThingSpeak for real-time monitoring and control via a mobile app. Fuzzy logic defines rules for efficient water distribution [13]. Additionally, an automated greenhouse irrigation system using an Arduino MEGA 2560 board demonstrates the effectiveness of IoT in maintaining optimal growing conditions by continuously adjusting irrigation based on sensor data [14]. These advancements illustrate the transformative potential of IoT and AI in achieving sustainable and efficient agricultural practices.

Artificial intelligence techniques, particularly computer vision, have shown significant potential in automating plant disease detection, enabling early and accurate identification of diseases. Deep learning-based computer vision methods, such as convolutional neural networks (CNNs), are increasingly used for the detection and classification of plant [15], [16]. These technologies allow for disease identification through the analysis of plant images, providing a more efficient and less labor-intensive alternative to manual monitoring [17]. Recent diseases advancements in computer vision and deep learning have facilitated the autonomous detection of plant diseases through the analysis of images captured by optical sensors, allowing for timely diagnosis of crop diseases [18]. Furthermore, the use of computer vision techniques in combination with AI has facilitated the early detection of plant diseases, allowing for timely interventions to mitigate the adverse effects of diseases [19].

Vision transformers (ViTs) have emerged as a significant advancement in the field of computer vision, building on the success of transformer models from natural language processing (NLP) [20]. These transformers, such ViTs, have demonstrated impressive performance across various machine vision tasks [21]. ViTs showcased their ability to achieve excellent results compared to state-of-the-art convolutional networks while requiring fewer computational resources for training [22]. Furthermore, ViTs have been applied to a wide range of computer vision applications, highlighting their versatility and potential [23]. Vision transformers represent a significant development in computer vision, providing a promising alternative to traditional convolutional neural networks. Researchers are continuously exploring and enhancing the capabilities of ViTs through studies focused on robustness, generalization, efficiency, and diverse applications, paving the way for further advancements in the field of computer vision.

With the advancements in deep learning, particularly the emergence of ViTs, there has been a significant shift towards automating this process. We explored recent articles applying ViTs in plant disease detection. A smartphone-based solution employing ViT models is proposed for identifying healthy and diseased tomato plants. The ViT model, trained on a dataset of tomato leaf images, outperforms traditional CNN-based approaches, demonstrating its potential for widespread adoption in smart agriculture systems [24]. Borhani *et al.* [25] explores ViTs for real-time automated plant disease classification. The study compares ViT with traditional CNN methods, highlighting the trade-offs between accuracy and prediction speed. It suggests potential enhancements through the combination of attention blocks with CNN blocks. In a different approach, authors introduce a fine-tuned technique called GreenViT for detecting plant infections and diseases. By leveraging ViTs, GreenViT overcomes the limitations associated with CNN-based models, demonstrating superior performance in detecting plant diseases [26]. Addressing the need for enhanced feature extraction, researchers proposes an edge-feature guidance module (EFG) to improve the feature extraction capabilities of ViT-based methods, leading to improved performance across multiple datasets [27]. For cassava leaf disease detection, ViT was used with techniques such as least important attention pruning (LeIAP) and sparse matrix-matrix multiplication (SPMM), resulting in significant improvements in accuracy and efficiency [28]. The study on plant disease classification presents a novel approach that integrates transfer learning with ViTs. This hybrid model achieves impressive validation accuracy, surpassing traditional transfer learning-based models. The efficiency of ViTs in extracting deep features from plant leaves is highlighted as a key factor in the model's superior performance [29]. In summary, the reviewed literature highlights the growing interest in leveraging ViTs for plant disease detection and classification. These studies contribute to advancing precision agriculture by providing efficient and accurate solutions for automated disease identification. Further research in this area could explore optimization techniques, model interpretability, and real-world deployment scenarios to enhance the practical applicability of ViTs in agricultural systems.

The aim of this paper is to explore ViTs in plant disease detection using a dataset containing different types of plants, in order to a further implementation in a smart agricultural system. The paper is structured into several main sections: a detailed methodology section, the presentation and discussion of results, and a conclusive summary. The methodology section outlines the experimental approach employed in this study, including dataset, data preprocessing and proposed model. Subsequently, the results and discussion section present the outcomes of the experiments, analyzing the performance of ViTs in plant disease detection tasks and discussing their implications for agricultural practices. Finally, the conclusion synthesizes the main findings, discusses their broader implications, and suggests avenues for future research. Through this structured approach, the paper aims to contribute to the advancement of plant disease detection methods and the promotion of sustainable agricultural practices.

2. METHOD

In this methodological section, we present the approaches and tools utilized to conduct our study. We begin by introducing the central dataset that forms the basis of our analyses, detailing its composition and preprocessing methods. Subsequently, we delve into an in-depth exploration of the innovative ViT architecture, a significant advancement in computer vision. The ViT distinguishes itself through its ability to effectively capture long-range dependencies in image data using self-attention mechanisms, thereby offering promising avenues for feature extraction and pattern recognition. This methodological introduction sets the stage for understanding the analyses and findings presented in this paper.

2.1. Proposed solution

To overcome the limitations of traditional plant disease detection methods, this study proposes the use of ViTs. ViTs leverage self-attention mechanisms to capture intricate patterns and long-range dependencies in plant images, offering a robust alternative to convolutional neural networks (CNNs). The methodology involves training a ViT model on a dataset of diverse plant images categorized by disease type. By partitioning images into patches and applying self-attention mechanisms, ViTs can effectively learn complex features and improve classification accuracy. The proposed solution integrates ViTs with advanced data preprocessing and augmentation techniques to enhance model performance and generalization across different plant species and disease conditions.

2.2. Dataset

The dataset from Kaggle consists of images of plant leaves categorized into 88 classes [30], representing various plant species and their health conditions. The dataset used in this study covers an extensive array of 55 classes from the original dataset, representing a substantial number of 14 plant species with 83.603 images. Figure 1 presents a snapshot of random samples from the dataset. The dataset utilized in this paper was extracted from the original database, and the images were augmented to achieve a balanced distribution across all categories. The new dataset encompasses a wide range of plants: apple, cassava, cherry, chili, corn, cucumber, grape, pomegranate, potato, soybean, strawberry, sugarcane and tomato. Within each plant category, different classes denote specific diseases or health conditions Table 1, resulting in a diverse collection of comprehensive machine learning model training.

2.3. Data preprocessing

As image preprocessing is a crucial step in preparing data for machine learning models, particularly in computer vision tasks. The process often involves augmenting the dataset to enhance the diversity and quantity of training samples, which helps improve the robustness and performance of the models. The augmenter defined here employs several techniques using the image library. It includes horizontal flipping ($iaa.Fliplr(0.5)$), which reverses images horizontally with a probability of 50%, and cropping ($iaa.Crop(percent = (0, 0.1))$), which randomly removes up to 10% of the image's borders. Contrast adjustments ($iaa.LinearContrast(0.75, 1.5)$) dynamically alter the image contrast, while additive Gaussian noise ($iaa.AdditiveGaussianNoise(scale = (0, 0.05 * 255))$) introduces slight randomness to pixel values to simulate real-world variations. Brightness changes ($iaa.Multiply(0.8, 1.2)$) adjust the image's brightness, making the model resilient to lighting conditions. Finally, affine transformations ($iaa.Affine(rotate = (-5, 5), shear = (-16, 16))$) involve rotating the image within a range of -5 to 5 degrees and shearing it between -16 and 16 degrees, effectively distorting the image while preserving its essential features. These augmentations collectively ensure that the dataset is varied and comprehensive, which is vital for training effective and generalized models. The distribution of images in each class of the new dataset is shown in Figure 2.

Figure 1. Sample of the dataset

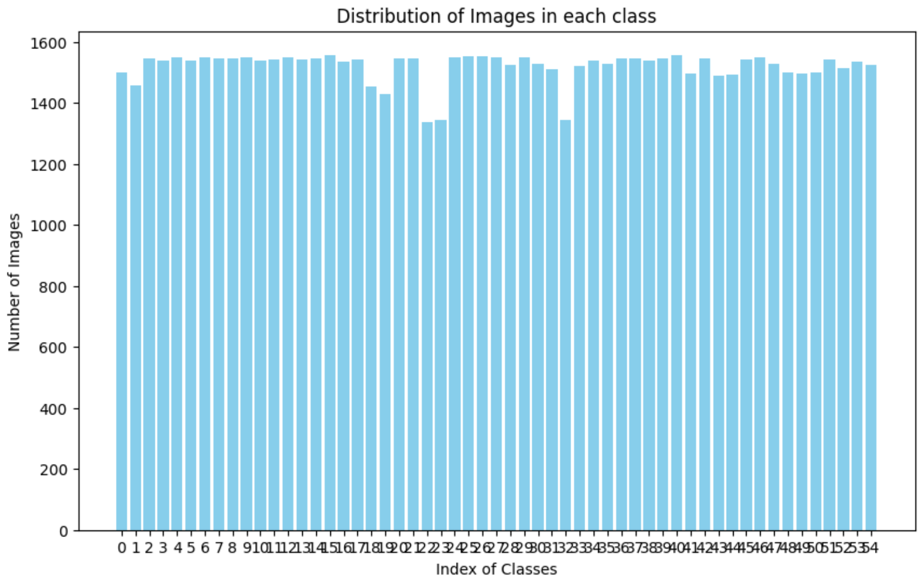

Figure 2. Distribution of images in each class

2.4. Vision transformers

The ViT architecture represents a significant advancement in the field of computer vision, leveraging the success of the transformer model in natural language processing tasks [22]. ViT has demonstrated remarkable performance in image classification, even surpassing traditional architectures like

ResNets [31]. Inspired by ViT, researchers have developed variations such as the swin transformer, which adapts the ResNet-50 architecture to create a hierarchical ViT [32]. These adaptations aim to enhance the original ViT design by integrating more recent training techniques without introducing additional attention-based modules. ViTs have gained popularity due to their success in various vision tasks, leading to the emergence of novel architectures like convolutional vision transformers (CvT) [33]. CvT combines the strengths of convolutions and Transformers to enhance performance and efficiency. Additionally, ViT has been explored in different domains beyond image classification, such as dense prediction tasks [34]. Overall, the ViT architecture signifies a pivotal shift in computer vision, showcasing its versatility and effectiveness across a wide range of applications.

Table 1. Descriptive of the plant and diseases included in the dataset

Plant	Diseases
Apple	Black rot, rust, scab, healthy
Cassava	Bacterial blight, brown streak disease, green mottle, healthy, mosaic disease
Cherry	Healthy, powdery mildew
Chili	Healthy, leaf curl, leaf spot, whitefly, yellowish
Corn	Common rust, gray leaf spot, healthy, northern leaf blight
Cucumber	Diseased, healthy
Grape	Black measles, black rot, healthy, leaf blight (isariopsis leaf spot)
Pomegranate	Diseased, healthy
Potato	Early blight, healthy, late blight
Soybean	Caterpillar, diabrotica speciosa, healthy
Strawberry	Healthy, leaf scorch
Sugarcane	Bacterial blight, healthy, red rot, red stripe, rust
Tomato	Bacterial spot, early blight, healthy, late blight, leaf mold, mosaic virus, septoria leaf spot, spider mites (two spotted spider mite), target spot, yellow leaf curl virus
Wheat	Brown rust, healthy, septoria, yellow rust

The ViT model is tailored for visual tasks like image classification, diverges from traditional CNNs by dividing input images into fixed-size patches, each transformed into a lower-dimensional vector space. These patch embeddings then feed into a stack of Transformer encoder layers. Within each encoder layer, two main sub-modules operate: a multi-head self-attention mechanism to capture long-range dependencies and a position-wise fully connected feedforward neural networks for context-aware representations. To address the lack of inherent sequence order to understand in Transformers, positional encodings are added to convey spatial information. Finally, a classification head, often a linear layer with SoftMax activation, is appended to the output for generating class predictions. This architecture's key hyperparameter is the dimensionality of patch embeddings, crucial for balancing model capacity and computational efficiency.

We present the proposed system in Figure 3. The system was developed using the database of plant disease images. The dataset was systematically divided into training, validation, and test subsets, with proportions of 80%, 10%, and 10%, respectively. A model was then created, trained, and validated using the training and validation subsets. Following this, the model's performance was rigorously tested using the test subset. The ultimate goal of this work was to develop a model capable of accurately predicting the class of plant diseases from images, thereby providing a valuable tool for agricultural diagnostics and management.

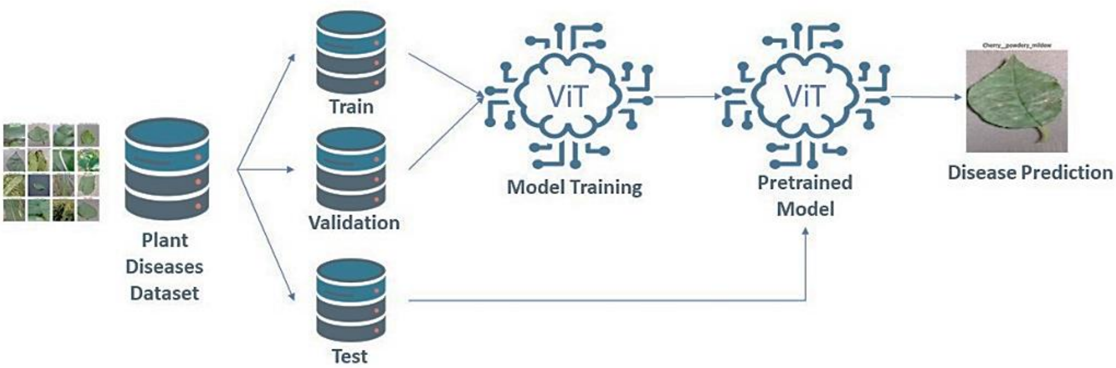

Figure 3. Proposed ViT system for plant disease detection

2.5. Evaluation metrics

The primary evaluation metric of our model is the F1-score, which is the harmonic mean of precision and recall. The F1-score is calculated as follows:

$$\text{F1 score} = \frac{2 * (precision * recall)}{(precision + recall)} \text{ [35]}$$

$$\text{Precision} = \frac{TP}{TP+FP} \text{ [36]}$$

$$\text{Recall} = \frac{TP}{TP+FN} \text{ [37]}$$

The terms TP, FP, and FN stand for: true positive (TP): The number of correct positive predictions. It refers to instances where the model correctly predicts the positive class. False positive (FP): The number of incorrect positive predictions. It refers to instances where the model incorrectly predicts the positive class, when the actual class is negative. False negative (FN): The number of incorrect negative predictions. It refers to instances where the model incorrectly predicts the negative class, when the actual class is positive.

3. RESULTS AND DISCUSSION

The ViT model presented in Table 2 implements a pioneering architecture for image-based plant diseases classification tasks, leveraging both patch-based encoding and transformer layers. Beginning with the PatchEncoder layer, input images are partitioned into patches, typically in size 16×16 pixels, extracted using a sliding window approach. Each patch undergoes a linear projection followed by positional embeddings, embedding spatial information into the data. This process creates a sequence of patch embeddings. The model architecture then integrates multiple layers of TransformerEncoder, each comprising multi-head self-attention mechanisms and position-wise feedforward networks. These transformer layers are pivotal in capturing both local and global dependencies within the image, facilitated by techniques like layer normalization and residual connections. Furthermore, the model incorporates configurable parameters such as the number of transformer heads, hidden size, and the number of patches, enabling flexibility and scalability. During training, the model's parameters are optimized using the Adam optimizer with a learning rate of 0.0001, ensuring efficient convergence. The encoded features are then flattened and processed through several dense layers, enhancing the model's capacity for learning intricate patterns. Finally, the output layer employs SoftMax activation to produce predictions for a predefined number of output classes, enabling the model to classify input images accurately. Through meticulous training with labeled image data and parameter tuning, the ViT model demonstrates exceptional performance in image classification tasks, showcasing its adaptability and efficacy across diverse visual recognition domains.

Table 2. Vision transformer model summary

Layer (type)	Output Shape	Param #
Input	(None, 256, 256, 3)	0
PatchEncoder	(None, 256, 512)	524,800
TransformerEncoder	(None, 256, 512)	8,665,088
TransformerEncoder	(None, 256, 512)	8,665,088
TransformerEncoder	(None, 256, 512)	8,665,088
TransformerEncoder	(None, 256, 512)	8,665,088
Dense	(None, 256, 256)	33,554,688
Dense	(None, 256, 2048)	526,336
Dense	(None, 256, 1024)	2,098,176
Dense	(None, 256, 512)	524,800
Dense	(None, 256, 128)	131,328
Dense	(None, 256, 64)	32,896
Dense	(None, 256, 32)	8,256
Dense	(None, 256, 55)	2,080
Output	(None, 55)	1,815

The model presented is a ViT architecture designed for image classification tasks. It consists of a PatchEncoder module that extracts image patches and encodes them using linear projections and positional embedding. These patches are then sequentially processed by a TransformerEncoder module, which applies multi-head self-attention and feedforward neural networks. The architecture is encapsulated within the ViT model, which includes additional dense layers before the final SoftMax output. Hyperparameters include eight heads, a hidden size of 512, 256 patches, four transformer layers, and 256 dense units. Trained with Adam optimizer (learning rate of 0.0001) and sparse categorical cross entropy loss, the model undergoes 20 epochs with a batch size of 32, achieving accuracy evaluation on test data. This model architecture demonstrates the effectiveness of transformer-based approaches in image classification tasks. The training process of the model exhibited a consistent improvement in accuracy, Figure 4 shows that the training process exhibited a consistent improvement in accuracy over 20 epochs, with accuracy steadily rising from approximately 24% to nearly 94.5%. In contrast, the validation accuracy generally lagged behind the training accuracy, with values ranging from approximately 44.7% to 91.6%. Figure 5 illustrates that the training loss decreased steadily from over 3.2 to around 0.13, indicating a progressive refinement in the model's performance. Similarly, the validation loss followed a decreasing trend, decreasing from over 1.78 to around 0.32, indicating that the model's generalization to unseen data improved over the training epochs. Results of Training accuracy and loss, validation accuracy and loss, and test accuracy and test are summarized in Table 3.

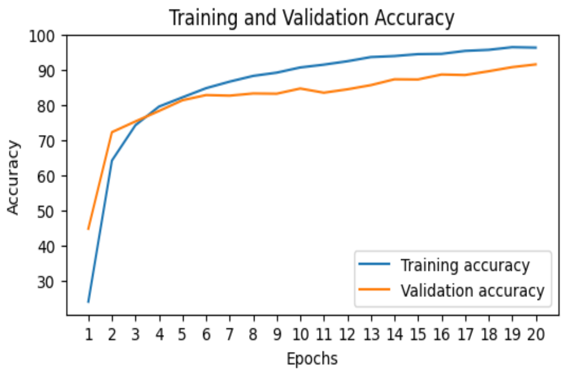

Figure 4. Training and validation accuracy

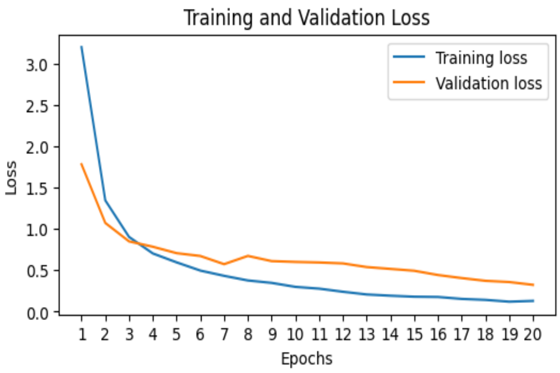

Figure 5. Training and validation loss

Table 3. Training accuracy and loss, validation accuracy and loss, and test accuracy and test

	Training	Validation	Test
Accuracy	94.5%	91.6%	89.3%
Loss	0.13	0.32	0.28

Table 4 presents the model evaluation metrics including precision, recall, and F1-score for various classes of plant diseases across different crops. Each class represents a specific disease or health condition, along with its corresponding evaluation metrics and support count. The table concludes with overall accuracy metrics for the model, along with macro and weighted averages across all classes. The ViT plant disease detection model demonstrates a strong overall performance with an accuracy of 90%, supported by macro and weighted average metrics around 90% for precision, recall, and F1-score, indicating consistent effectiveness across various plant classes and conditions. For specific classes, the model excels in detecting apple diseases such as black rot, rust, and scab, with F1-scores ranging from 0.88 to 0.93, and identifies healthy apple conditions with a 0.90 F1-score. In cassava, it shows perfect detection for mosaic disease and high precision for bacterial blight and brown streak disease, though it struggles slightly with green mottle, which has a lower recall (0.79) and an F1-score of 0.88. For cherry, the model achieves high F1-scores for both healthy (0.91) and powdery mildew (0.92) conditions. Chili diseases are well-detected, particularly leaf curl and healthy conditions, with F1-scores of 0.90-0.92, although whitefly detection has a lower precision (0.76) but a high recall (0.94), resulting in an F1-score of 0.84. Corn disease detection varies, with common rust having the lowest F1-score (0.81) due to lower precision, while gray leaf spot and northern leaf blight show very high F1-scores of 0.90 and 0.96, respectively. The model also performs well for cucumber, grape, and pomegranate diseases, achieving perfect or near-perfect scores for several conditions. In potato, early blight is detected with a lower recall (0.77) but maintains a reasonable F1-score (0.85). These results highlight the model's robustness and effectiveness in identifying a wide range of plant diseases.

Table 4. Model evaluation metrics (Precision, Recall, F1-Score)

	Class	Precision	Recall	F1-Score	Support
Apple	Black rot	0.83	0.94	0.88	200
	Rust	0.93	0.93	0.93	200
	Scab	0.90	0.86	0.88	200
	Healthy	0.86	0.95	0.90	200
Cassava	Bacterial blight	0.96	0.92	0.94	200
	Brown streak disease	0.90	0.95	0.93	200
	Green mottle	1.00	0.79	0.88	200
	Healthy	0.90	0.86	0.88	200
	Mosaic disease	1.00	1.00	1.00	200
Cherry	Healthy	0.95	0.87	0.91	200
	Powdery mildew	0.89	0.94	0.92	200
Chili	Healthy	0.81	1.00	0.90	200
	Leaf curl	0.86	1.00	0.92	200
	Leaf spot	0.92	0.96	0.94	200
	Whitefly	0.76	0.94	0.84	200
	Yellowish	0.91	0.95	0.93	200
Corn	Common rust	0.73	0.92	0.81	200
	Gray leaf spot	0.83	1.00	0.90	200
	Healthy	0.93	0.88	0.90	200
	Northern leaf blight	0.96	0.96	0.96	200
Cucumber	Diseased	0.83	0.90	0.86	200
	Healthy	0.94	0.89	0.92	200
Grape	Black measles	0.86	0.92	0.89	200
	Black rot	0.90	1.00	0.95	200
	Healthy	1.00	0.93	0.97	200
	Leaf blight (isariopsis leaf spot)	0.93	0.93	0.93	200
Pomegranate	Diseased	1.00	0.88	0.93	200
	Healthy	0.96	0.92	0.94	200
Potato	Early blight	0.94	0.77	0.85	200
	Healthy	0.94	0.89	0.91	200
	Late blight	0.89	0.89	0.89	200
Soybean	Caterpillar	0.89	0.94	0.91	200
	Diabrotica speciosa	0.95	0.83	0.89	200
	Healthy	0.94	0.88	0.91	200
Strawberry	Healthy	0.92	0.85	0.88	200
	Leaf scorch	0.81	0.94	0.87	200
Sugarcane	Bacterial blight	0.89	0.89	0.89	200
	Healthy	1.00	1.00	1.00	200
	Red rot	0.86	0.86	0.86	200
	Red stripe	0.86	0.92	0.89	200
	Rust	1.00	0.88	0.93	200
Tomato	Bacterial spot	0.95	0.90	0.92	200
	Early blight	0.92	0.92	0.92	200
	Healthy	0.88	0.82	0.85	200
	Late blight	0.85	0.96	0.90	200
	Leaf mold	0.92	0.86	0.89	200
	Mosaic virus	0.85	1.00	0.92	200
	Septoria leaf spot	0.94	0.83	0.88	200
	Spider mites (two spotted spider mite)	0.86	0.86	0.86	200
	Target spot	1.00	0.87	0.93	200
	Yellow leaf curl virus	1.00	0.94	0.97	200
Wheat	Brown rust	0.94	0.89	0.92	200
	Healthy	0.95	0.95	0.95	200
	Septoria	0.67	0.71	0.69	200
	Yellow rust	0.81	0.76	0.79	200
Accuracy		0.90	0.90	0.90	11000
macro avg		0.90	0.91	0.90	11000
weighted avg		0.91	0.90	0.90	11000

4. CONCLUSION

Plant diseases continue to pose significant challenges to global agriculture, threatening food security and economic stability. Traditional methods for disease detection are often labor-intensive, subjective, and limited in scalability, prompting the need for more efficient and accurate approaches. This paper has reviewed the growing interest in ViTs for automated plant disease detection, showcasing their potential to revolutionize agricultural practices. Through a comprehensive methodology and experimental evaluation, ViTs have demonstrated exceptional performance in classifying diverse plant diseases across multiple datasets. These results underscore the effectiveness of ViTs in capturing complex patterns in plant images, enabling accurate and timely disease identification. The dataset preprocessing has shown a significant role in model accuracy as the classes were balanced to improve the model. The adoption of ViTs in precision agriculture holds promise for enhancing crop productivity, minimizing losses, and promoting sustainable farming practices. Future research directions may focus on optimizing ViT architectures, improving interpretability, and exploring real-world deployment scenarios to facilitate the widespread adoption of these technologies in agricultural systems. Overall, ViTs represent a significant advancement in computer vision for agriculture, offering transformative solutions to mitigate the impact of plant diseases and promote global food security.

ACKNOWLEDGEMENTS

This work was held in the framework of the Mixed Morocco-Tunisia Laboratory "Laboratoire Morocco-Tunisia: Environnement et développement Durable (E2D)". It was supported by Ministry of National Education, Vocational Training, Higher Education and Scientific Research, Department of Higher Education and Scientific Research, in Morocco, and by Ministry of High Education and Scientific Research in Tunisia.

REFERENCES

[1] R. M. Nadeem, A. Jaffar, and R. M. Saleem, "IoT and machine learning based stem borer pest prediction," *Intelligent Automation and Soft Computing*, vol. 31, no. 3, pp. 1377–1392, 2022, doi: 10.32604/IASC.2022.020680.

[2] S. Sudheer, R. G. Bai, Z. Usmani, and M. Sharma, "Insights on engineered microbes in sustainable agriculture: biotechnological developments and future prospects," *Current Genomics*, vol. 21, no. 5, pp. 321–333, 2020, doi: 10.2174/1389202921999200603165934.

[3] A. Nadhiya, "Evolution of management consulting and its siginificance in the sustainable development of agrarian enterperises," *InterConf*, no. 30(143), pp. 32–40, 2023, doi: 10.51582/interconf.19-20.02.2023.005.

[4] R. M. J. Al-Akkam and M. S. M. Altaei, "Plants leaf diseases detection using deep learning," *Iraqi Journal of Science*, vol. 63, no. 2, pp. 801–816, 2022.

[5] P. Srivastava, K. Mishra, V. Awasthi, V. K. Sahu, and P. K. Pal, "Plant disease detection using convolutional neural network," *International Journal For Multidisciplinary Research*, vol. 6, no. 3, 2024, doi: 10.36948/ijfmr.2024.v06i03.19536.

[6] C. Usha Kumari, N. Arun Vignesh, A. K. Panigrahy, L. Ramya, and T. Padma, "Fungal disease in cotton leaf detection and classification using neural networks and support vector machine," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 3664–3667, 2019, doi: 10.35940/ijitee.J9648.0881019.

[7] R. Balodi, S. Bisht, A. Ghatak, and K. H. Rao, "Plant disease diagnosis: technological advancements and challenges," *Indian Phytopathology*, vol. 70, no. 3, pp. 275–281, 2017, doi: 10.24838/ip.2017.v70.i3.72487.

[8] A. J. Mastin, F. Van Den Bosch, F. Van Den Berg, and S. R. Parnell, "Quantifying the hidden costs of imperfect detection for early detection surveillance," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 374, no. 1776, 2019, doi: 10.1098/rstb.2018.0261.

[9] D. Migliorini, L. Ghelardini, E. Tondini, N. Luchi, and A. Santini, "The potential of symptomless potted plants for carrying invasive soilborne plant pathogens," *Diversity and Distributions*, vol. 21, no. 10, pp. 1218–1229, 2015, doi: 10.1111/ddi.12347.

[10] C. Temple, A. G. Blouin, S. Tindale, S. Steyer, K. Marechal, and S. Massart, "High throughput sequencing technologies complemented by grower's perception highlight the impact of tomato virome in diversified vegetable farms," *Frontiers in Sustainable Food Systems*, 2023, doi: 10.1101/2023.01.12.523758.

[11] S. Padhee and D. Nandan, "Design of automated visual inspection system for beverage industry production line," *Traitement du Signal*, vol. 38, no. 2, pp. 461–466, 2021, doi: 10.18280/ts.380225.

[12] A. Mhaned, M. Salma, E. H. Mounia, and B. Jamal, "Contribution to smart irrigation based on internet of things and artificial intelligence," *Lecture Notes in Networks and Systems*, vol. 625 LNNS, pp. 537–549, 2023, doi: 10.1007/978-3-031-28387-1_45.

[13] A. Mhaned, S. Mouatassim, M. El Haji, and J. Benhra, "Low-cost smart irrigation system based on internet of things and fuzzy logic," *Communications in Computer and Information Science*, vol. 1677 CCIS, pp. 78–89, 2022, doi: 10.1007/978-3-031-20490-6_7.

[14] S. Fouguira, A. Mhaned, M. Ben Abbou, E. Ammar, M. El Haji, and J. Benhra, "Effectiveness of compost use in salt-affected soil in an automated greenhouse irrigation system," *E3S Web of Conferences*, vol. 364, 2023, doi: 10.1051/e3sconf/202336403002.

[15] R. Sharma *et al.*, "Plant disease diagnosis and image classification using deep learning," *Computers, Materials and Continua*, vol. 71, no. 2, pp. 2125–2140, 2022, doi: 10.32604/cmc.2022.020017.

[16] H. T. Rauf, B. A. Saleem, M. I. U. Lali, M. A. Khan, M. Sharif, and S. A. C. Bukhari, "A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning," *Data in Brief*, vol. 26, 2019, doi: 10.1016/j.dib.2019.104340.

[17] M. E. H. Chowdhury *et al.*, "Automatic and reliable leaf disease detection using deep learning techniques," *AgriEngineering*, vol. 3, no. 2, pp. 294–312, 2021, doi: 10.3390/agriengineering3020020.

[18] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Detection of apple lesions in orchards based on deep learning methods of cyclegan and YoloV3-dense," *Journal of Sensors*, vol. 2019, 2019, doi: 10.1155/2019/7630926.

[19] K. J. Ayikpa, D. Mamadou, P. Gouton, and K. J. Adou, "Experimental evaluation of coffee leaf disease classification and recognition based on machine learning and deep learning algorithms," *Journal of Computer Science*, vol. 18, no. 12, pp. 1201–1212, 2022, doi: 10.3844/jcssp.2022.1201.1212.

[20] T. Li, F. Zhang, G. Xie, X. Fan, Y. Gao, and M. Sun, "A high speed reconfigurable architecture for softmax and GELU in vision transformer," *Electronics Letters*, vol. 59, no. 5, 2023, doi: 10.1049/ell2.12751.

[21] M. Naseer, K. Ranasinghe, S. Khan, M. Hayat, F. S. Khan, and M. H. Yang, "Intriguing properties of vision transformers,"

Advances in Neural Information Processing Systems, vol. 28, pp. 23296–23308, 2021.
[22] A. Dosovitskiy *et al.*, "An image is worth 16X16 words: transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2021.
[23] S. Jamil, M. Jalil Piran, and O. J. Kwon, "A comprehensive survey of transformers for computer vision," *Drones*, vol. 7, no. 5, 2023, doi: 10.3390/drones7050287.
[24] U. Barman *et al.*, "ViT-SmartAgri: vision transformer and smartphone-based plant disease detection for smart agriculture," *Agronomy*, vol. 14, no. 2, 2024, doi: 10.3390/agronomy14020327.
[25] Y. Borhani, J. Khoramdel, and E. Najafi, "A deep learning based approach for automated plant disease classification using vision transformer," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-15163-0.
[26] S. Parez, N. Dilshad, N. S. Alghamdi, T. M. Alanazi, and J. W. Lee, "Visual intelligence in precision agriculture: exploring plant disease detection via efficient vision transformers," *Sensors*, vol. 23, no. 15, 2023, doi: 10.3390/s23156949.
[27] B. Chang, Y. Wang, X. Zhao, G. Li, and P. Yuan, "A general-purpose edge-feature guidance module to enhance vision transformers for plant disease identification," *Expert Systems with Applications*, vol. 237, 2024, doi: 10.1016/j.eswa.2023.121638.
[28] H. T. Thai, K. H. Le, and N. L. T. Nguyen, "FormerLeaf: an efficient vision transformer for Cassava Leaf Disease detection," *Computers and Electronics in Agriculture*, vol. 204, 2023, doi: 10.1016/j.compag.2022.107518.
[29] A. Tabbakh and S. S. Barpanda, "A deep features extraction model based on the transfer learning model and vision transformer 'TLMViT' for plant disease classification," *IEEE Access*, vol. 11, pp. 45377–45392, 2023, doi: 10.1109/ACCESS.2023.3273317.
[30] A. Dobrovsky, "Plant disease classification merged dataset," *kaggle.com*, [Online]. Available: https://www.kaggle.com/datasets/alinedobrovsky/plant-disease-classification-merged-dataset (accessed Jun 21, 2024).
[31] S. Bhojanapalli, A. Chakrabarti, D. Glasner, D. Li, T. Unterthiner, and A. Veit, "Understanding robustness of transformers for image classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 10211–10221, doi: 10.1109/ICCV48922.2021.01007.
[32] M. Cantone, C. Marrocco, F. Tortorella, and A. Bria, "Convolutional networks and transformers for mammography classification: an experimental study," *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031229.
[33] H. Wu *et al.*, "CvT: introducing convolutions to Vision transformers," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 22–31, doi: 10.1109/ICCV48922.2021.00009.
[34] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense pediction," in *Proceedings of the IEEE International Conference on Computer Vision*, 2021, pp. 12159–12168, doi: 10.1109/ICCV48922.2021.01196.
[35] S. Ramesh and D. Vydeki, "Recognition and classification of paddy leaf diseases using optimized deep neural network with Jaya algorithm," *Information Processing in Agriculture*, vol. 7, no. 2, pp. 249–260, 2020, doi: 10.1016/j.inpa.2019.09.002.
[36] N. Agrawal, J. Singhai, and D. K. Agarwal, "Grape leaf disease detection and classification using multi-class support vector machine," in *International Conference on Recent Innovations in Signal Processing and Embedded Systems, RISE 2017*, 2018, vol. 2018-Janua, pp. 238–244, doi: 10.1109/RISE.2017.8378160.
[37] A. Parikh, M. S. Raval, C. Parmar, and S. Chaudhary, "Disease detection and severity estimation in cotton plant from unconstrained images," in *Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016*, 2016, pp. 594–601, doi: 10.1109/DSAA.2016.81.

BIOGRAPHIES OF AUTHORS

Mhaned Ali holds a degree in industrial engineering from Mundiapolis University of Casablanca, completed in 2019. Prior to that, he earned a diploma in technological studies in electrical engineering from Casablanca Higher School of Technology in 2014, where he specialized in local industrial networks and industrial supervision. Currently, pursuing a Ph.D. in smart agricultural systems with a focus on IoT and artificial intelligence at the National High School of Electrical and Mechanical Engineering. His research is conducted at the Laboratory of Advanced Research in Industrial and Logistic Engineering (OSIL Team), where he applies his skills in IoT, AI, machine learning, and deep learning to develop innovative irrigation solutions. He can be contacted at email: ali.mhaned.doc21@ensem.ac.ma.

Mouatassim Salma is a professor and researcher at ENSEM. She holds a Ph.D. in engineering sciences with a specialization in industrial engineering, an engineering degree from ENSEM, and a research master's degree in industrial systems engineering from ENIM, University of Lorraine. Her research is centered on the optimization of end-to-end supply chains, with a particular focus on integrating artificial intelligence and advanced data analytics. Her work aims to enhance efficiency, reduce costs, and improve decision-making across various stages of the supply chain, from procurement to distribution. She can be contacted at email: s.mouatassim@ensem.ac.ma.

Mounia El Haji a professor of higher education at the National High School of Electrical and Mechanical Engineering in Casablanca, where she has been teaching Industrial Engineering and Logistics since September 2015. She is part of the Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE) and the Research Team Optimization of Industrial and Logistical Systems (OSIL). She holds a Doctorate in Physico-Chemistry and Water Quality from the University of Le Havre 1998 and a DEA National in Hydrology and Hydrogeology from Université Paris 6 (Jussieu) 1995. She can be contacted at email: m.elhaji@ensem.ac.ma.

Benhra Jamal currently professor in the Department of Industrial and logistic Engineering, in ENSEM, Casablanca. He is the Head of the Laboratory of Advanced Research in Industrial and Logistic Engineering (LARILE). He does research in optimization algorithms, artificial intelligence and computing in mathematics, natural science, engineering and medicine. He can be contacted at email: jamalbenhra@gmail.com.