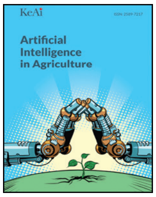
Contents lists available at ScienceDirect

Artificial Intelligence in Agriculture

journal homepage: http://www.keaipublishing.com/en/journals/artificial-intelligence-in-agriculture/



CactiViT: Image-based smartphone application and transformer network for diagnosis of cactus cochineal

Anas Berka [a,b,*], Adel Hafiane [b], Youssef Es-Saady [a], Mohamed El Hajji [a], Raphaël Canals [b], Rachid Bouharroud [c]

[a] IRF-SIC Laboratory, Ibnou Zohr University, BP 8106—Cite Dakhla, Agadir 80000, Morocco
[b] INSA CVL, University of Orleans, PRISME Laboratory, EA 4229, Bourges 18022, France
[c] Regional Center for Agricultural Research of Agadir, National Institute of Agricultural Research, Avenue Ennasr, PoB 415 Rabat Principale, Rabat 10090, Morocco

ARTICLE INFO

Article history:
Received 24 August 2022
Received in revised form 14 June 2023
Accepted 5 July 2023
Available online 13 July 2023

Keywords:
Cactus
Cochineal
Smartphones
Classification VIT
Deep learning

ABSTRACT

The cactus is a plant that grows in many rural areas, widely used as a hedge, and has multiple benefits through the manufacture of various cosmetics and other products. However, this crop has been suffering for some time from the attack of the carmine scale *Dactylopius opuntia* (Hemiptera: Dactylopiidae). The infestation can spread rapidly if not treated in the early stage. Current solutions consist of regular field checks by the naked eyes carried out by experts. The major difficulty is the lack of experts to check all fields, especially in remote areas. In addition, this requires time and resources. Hence the need for a system that can categorize the health level of cacti remotely. To date, deep learning models used to categorize plant diseases from images have not addressed the mealy bug infestation of cacti because computer vision has not sufficiently addressed this disease. Since there is no public dataset and smartphones are commonly used as tools to take pictures, it might then be conceivable for farmers to use them to categorize the infection level of their crops. In this work, we developed a system called CactiVIT that instantly determines the health status of cacti using the Visual image Transformer (ViT) model. We also provided a new image dataset of cochineal infested cacti.[1] Finally, we developed a mobile application that delivers the classification results directly to farmers about the infestation in their fields by showing the probabilities related to each class. This study compares the existing models on the new dataset and presents the results obtained. The VIT-B-16 model reveals an approved performance in the literature and in our experiments, in which it achieved 88.73% overall accuracy with an average of +2.61% compared to other convolutional neural network (CNN) models that we evaluated under similar conditions.

1. Introduction

Cactus is a plant that is used in many communities around the world as a land boundary or crop. It plays a fundamental role in improving land protection against erosion while contributing to the preservation of natural resources and biodiversity (Le Houérou, 1996). It is used for various purposes such as food, medicine, and cosmetics (Stintzing and Carle, 2005). In addition, in several countries, the cactus is considered an alternative crop with a high added value product for the development of arid and semi-arid areas (Griffith, 2004). Most of their cultivars are found in North America (Tigano et al., 2020). Other countries such as Algeria, Argentina, Ethiopia, France, Italy, Mexico, Morocco, South Africa, and Tunisia also cultivate it (Amani et al., 2019).

Despite efforts to protect and maximize cactus production, it is threatened by several diseases and pests. These include *Phyllosticta opuntiae* rust, *Phytophthora cactorum* cactus downy mildew, *Ceratitis capitata* ceratitis and *Dactylopius opuntiae* (Cockerell) opuntia cochineal scale (Nobel, 2002; Dodd et al., 1940; Donkin, 1977). Cochineal infestation is the main current problem for the cactus; it was first detected in Morocco in late 2014 (Bouharroud et al., 2016) and quickly spread to several provinces. There have also been other reports of this pest attack from other countries (Miller, 1996; Foldi, 2001; Aldama et al., 2005; Spodek et al., 2014). As for North and South America, the insect was initially used to control the invasive spread of the opuntia (Lotto, 1974; Foxcroft and Hoffmann, 2000): this pest causes severe damage to the plant by secreting a white wax on its leaves and fruits (Lotto, 1974), thus the visual indication in images showing that the cactus has been infected. For example, an early-stage infestation can be seen in Fig. 1.a while in Fig. 1.b, a late-stage infection with the cochineal is displayed. Early detection of diseases prevents damage to the crops in general (DiMiceli et al., 2021), so it is important for farmers to diagnose the

[*] Corresponding author.
E-mail address: anas.berka@insa-cvl.fr (A. Berka).

[1] The dataset is open-source and available at: https://github.com/AnasBerka/CactiViT-materials.git

https://doi.org/10.1016/j.aiia.2023.07.002

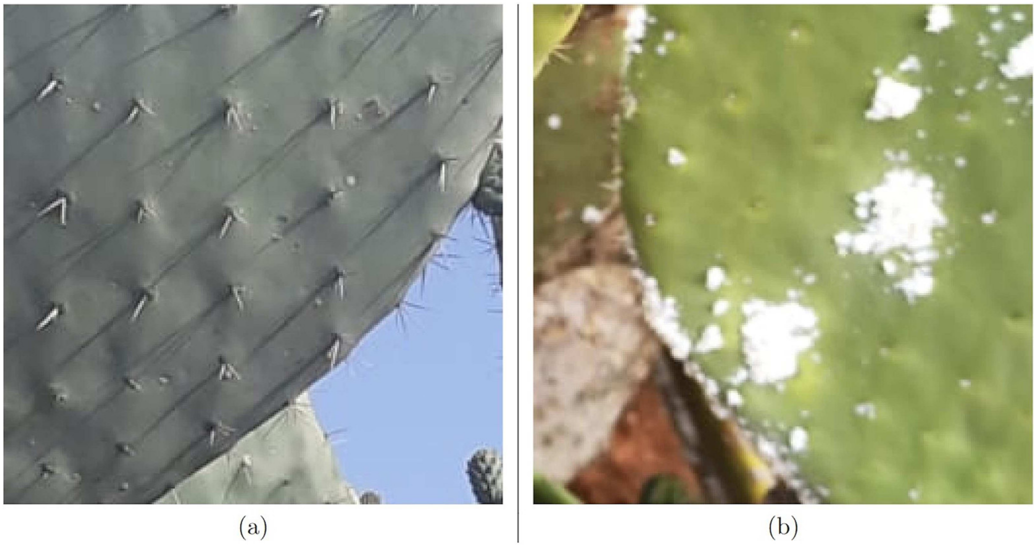


Fig. 1. Early-stage infection (a) and late-stage infection (b) of the cochineal on cactus.

disease. However, this requires regular field checks with naked eyes conducted by experts, which is generally not within their reach (Ramesh et al., 2018).

On the other hand, the development of artificial intelligence and mobile applications has opened up several opportunities for developing new technological approaches to provide valuable assistance tools to farmers (Li et al., 2020; Zhu et al., 2021). As smartphones are commonly used tools to acquire photos, they could be used to recognize plants, detect diseases, classify the health statuses and assess the infection level of a crops in the fields. Indeed, in recent years, more and more people have smartphones (Carton et al., 2018). As a result, several mobile applications have emerged in this field (Pongnumkul et al., 2015), for example such as Pl@ntNet, which provides the characteristics of a plant from a smartphone image using deep learning approaches (Goëau et al., 2013). Plantix is another diagnostic mobile app that uses plant images to detect diseases, pests, and nutritional deficiencies (GmbH, 2019; Icrisat, 2019). *E*-agree can detect, from a picture of the plant's leaves, the type of disease (Reddy et al., 2015). The applications we cited have a common type of processing that is a cloud-base computation in which the system is split into two, a cloud-based computation and a mobile server (Andrianto et al., 2020). Knowing that the internet is not always available, other apps offer local inference such as BioLeaf, in which the farmer can identify regions of injury on a leaf attacked by insects (Machado et al., 2016). Similarly, Plant Disease App targets grapevine diseases on images with more options for the user to achieve the highest accuracy (Petrellis, 2017).

Indeed, deep learning has enabled significant advances in plant disease identification through various approaches (Martinelli et al., 2015; Ferentinos, 2018; Golhani et al., 2018; Ngugi et al., 2021; Ouhami et al., 2021). Starting with AlexNet, the first neural network architecture to go beyond the support vector machine known as SVM (a machine learning model) (Krizhevsky et al., 2012), numerous models have been created and evaluated using the public PlantVillage dataset of 54.306 images (Hughes et al., 2015; Yuan et al., 2022). The focus has always been on proposing a better deep learning model to diagnose the diseases; models such as AlexNet combined with GoogleNet(Mohanty et al., 2016; Brahimi et al., 2017), SqueezeNet(Durmuş et al., 2017), or even ResNet almost full (Zhang et al., 2018). InceptionV3 also showed a comparable performance (Brahimi et al., 2018; Qiang et al., 2019). DenseNet121 scored nearly perfect (Too et al., 2019), but EfficientNet B4 and B5 (Atila et al., 2021) were the best performing models on PlantVillage, with 99.97% and 99.91% respectively. This raises many questions about the performance of deep learning on specific crops with a smaller image dataset. In fact, many researchers have proven the effectiveness of transfer learning using variety of models with transfer learning from The ImageNet dataset (Krizhevsky et al., 2017). These techniques have been tested on crops for which data are available, such as citrus (Shrivastava et al., 2019; Barman et al., 2020; Kaur et al., 2020), tomatoes (Fang and Ramasamy, 2015; Xie et al., 2017; El Massi et al., 2021; Ouhami et al., 2020), and grapevines (MacDonald et al., 2016; Wang et al., 2019; Kerkech et al., 2020; Reedha et al., 2022).

Recent developments in computer vision (Gao et al., 2021; Liang et al., 2021; Gheflati and Rivaz, 2022) show that the results from the Visual image Transformers (ViT) using the transformers architecture and the self-attention and multi-head mechanism in (Vaswani et al., 2017) are promising. (Dosovitskiy et al., 2020) achieve optimal accuracy using ViT-H and ViT-B on ImageNet. This proved that transformers can focus on the image, extract relevant information, and that the use of transformers cannot be limited to the natural language processing (NLP) problem, which is revolutionary (Wolf et al., 2020). Second, a higher accuracy of 90.45% on ImageNet classification was recorded after scaling the transformers architecture (Zhai et al., 2021). But for ViT to achieve this performances, pre-training on an exceptionally large dataset is required (Touvron et al., 2020). The use of transformers in agriculture has successfully detected diseases in vineyards (Reedha et al., 2022) and cassava (Thai et al., 2021). Within a few years, transformers will likely to be used for many more tasks in artificial intelligence. To the best of our knowledge, few researchers have addressed the cochineal infection problem from a computer vision perspective. In (Atitallah et al., 2021; Kaweesinsakul et al., 2021; Perez et al., 2022) the authors addressed the detection, classification of various cactus species, and segmentation of the plant cactus in complex environments. Since no public dataset is available, it is necessary to have a dataset of images acquired by smartphone for model training; therefore we have first constructed our image dataset.

In this paper, we propose an assistant system to diagnose the health condition of the cacti based on mobile images and Visual image Transformers (VIT). The system we propose is composed of two main components as described in Fig. 2: a remote server and a mobile application on edge device. The first one is for data collection and training the ViT models. The second one is operated on the mobile application, where a user can take a photo of cactus, obtain information concerning the health status of the cacti and provide a feedback. To avoid running a heavy computations on the smartphone, a remote machine capable of

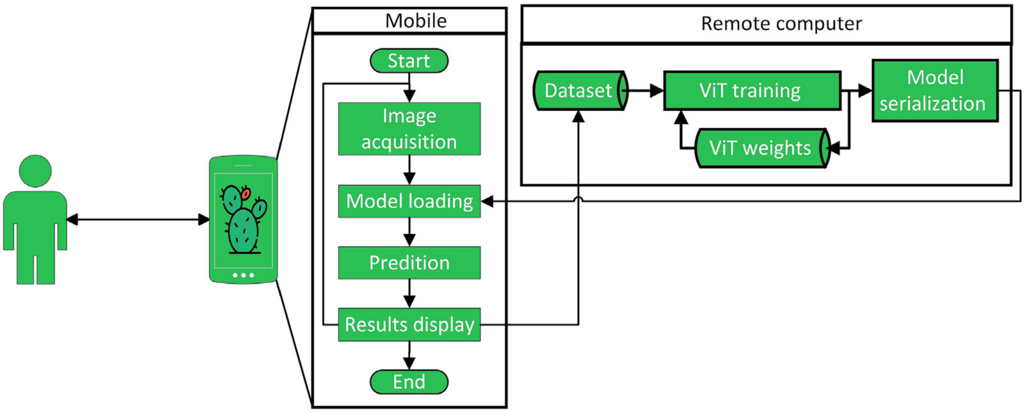


Fig. 2. System overview.

performing the storage and training tasks, was used to store the dataset and regularly re-train the ViT model to perform updates. The latest update will then be sent to the individual users' s edge device via the application for better accuracy. On the other side, the smartphone will acquire new cactus photos or select some from the gallery and perform the classification. We also contributed with new dataset for images of cochineal infested cacti. The proposed system could be a valuable tool for farmers growing cacti.

This work is organized as follows; the first section provides a brief overview of the research. The second section presents a description of the dataset, the materials and the methodology used. The third section details the conduct of experiences. In the fourth section, results are reported with a discussion in the same section. Then conclusions are drawn in the final section.

2. Materials and methods

2.1. Dataset description

The main dataset for this case study was created manually from the field using a smartphone equipped with an f/1.9 5 mm ISO-101 camera at several locations in Morocco, precisely in the Guelmim-Oued Noun region, on two dates: 27/03/2021 and 23/05/2021. On each date, we acquired data from different fields as shown in Table 1.

The raw images collected from the smartphone had many variations in saturation, angle of shot, resolution, range and zoom. After acquiring the images, they were resized into smaller images in the same shape $(256 \times 256 \times 3)$ without data augmentation. These resulting images were labeled using our manual classification process shown in Fig. 3 in order to perform a supervised learning approach. Our technique was to some extent based on (Vasconcelos et al., 2009; Akroud et al., 2021) since the authors had classified the infestation from 0 to 5 to differentiate between the various heath statuses. For comparison to our work, 0 is healthy, 1 is early stage and from 2 to 5 late stage. The principle is to look at the image and then answer the question to know what the class is for the given image. For example, if cacti are present in the image and there is a whitish sticky residue on them and there are many of them, then it is a late-stage infestation.

Table 1
Dataset acquisition.

Field N°	Name of the field	Acquisition date	Number of original images
1	Douar Tamgert Lmouden	27/03/2021	30
2	Douar Tigenda Mirleft	27/03/2021	14
3	Douar Tighratin Mirleft	27/03/2021	37
4	Route Taandit Mirleft	27/03/2021	42
5	Route Tigenda	27/03/2021	16
6	Douar Mrah	23/05/2021	141
7	Douar Mtguayzin	23/05/2021	147
8	Douar Taandit	23/05/2021	218

From each image acquired at a different location labeled as “Field i” where $i \in [1;8]$, we extracted several smaller images corresponding to the relevant classes. Following our method of labeling, we created the first dataset indicating the level of infestation by the carmine cochineal *Dactylopius opuntiae* in cacti. Manual classification with expert verification for each image should give one of the seven outputs shown in the diagram in Fig. 3. Note that the “Confused” class is a set of images with too much noise, a lot of mixed information or simply confusion where we cannot define the exact class just by looking at the image. This is because this class is a filter that eliminates any confusion that requires on-site checking to determine which class it belongs to. This class will not contribute to the learning or testing phase. It is stored in an isolated folder out of the experiences. By having the “Confused” class outside of the prepared dataset, we use 3921 images. In Table 2 we present the number of images generated in each class for the different locations. All the images were divided by the assigned class and not by the relative field of the source. The final verification was done with the help of the same expert to be sure of the correct labels for each image.

An example of each class is presented in Fig. 4. The difference between each class is the level of infestation and the health status of the cacti. In the Healthy class, there is both the spiny rackets and spineless rackets. In the Damaged class, we find any infestation other then the cochineal, while in the Old_Dead class, it is a class of dead cactus rackets. The classes NoCactus, EarlyStage and LateStage contain the most of the images. Our dataset is open-source and available at: https://github.com/AnasBerka/CactiViT-materials.git.

2.2. Model: Visual image transformers

The Visual image Transformer opted as the primary model for this study is practically the same as the one proposed by (Dosovitskiy et al., 2020) for the “ViT-Base”, as shown in Fig. 5. This model was originally based on the transformer of (Vaswani et al., 2017).

Let $S = \{X_i, y_i\}_{i=1}^{n}$ be a set of $n = 16$ images of the batch extracted from our prepared labeled dataset. The model takes as input i, an image X_i of size $(224 \times 224 \times 3)$ to predict its true label y_i. Were $y_i \in [1;6]$ corresponds to the encoded labels of the corresponding class in the data set. In the first step, at the embedding layer, the image X_i was reshaped from $(H \times W \times C)$ to $(N \times P \times P \times C)$ where $(H, W) = (224, 224)$ the resolution of the input image, $C = 3$ the number of channels, P the size of the patch used which can be 16 or 32 and $N = (H \times W)/(P \times P)$ the number of the new patches. N will be used as the number of effective input sequence lengths for the transformer. After that, a learning embedding matrix z is created to be used as an input in the next layer. z_0 is the first entry to the transformer encoder, the resulting patches as:

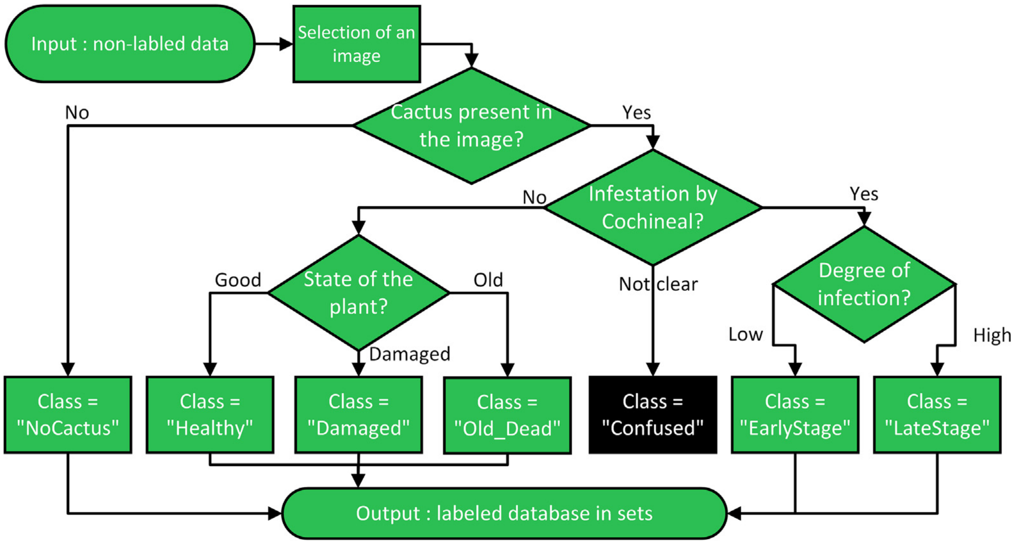


Fig. 3. Architecture of the data labeling used.

$$z_0 = \left[x_{class}; x_p^1 E; x_p^2 E; \cdots; x_p^N E\right] + E_{pos} \tag{1}$$

with $z_0^0 = x_{class}$ is a reserved token for the output of the transformer encoder layer, x_p^j is a patch of the input image, with $j \in 1, 2, ., N$ is the index of x_p on position p in X_i, $E \in R^{p^2 \times C}$ is the same learning embedding matrix to all patches and $E_{pos} \in R^{(N+1) \times D}$, where $D = P^2 \times C$, that adds position information for the patches. Afterwards the vector z_0 is forwarded to the transformer encoder with $L = 12$ layers for the "ViT-Base". The sequence z_l where $l = 1 \ldots 12$ is then updated as follows:

$$z'_l = MSA(LN(z_{l-1})) + z_{l-1} \tag{2}$$

$$z_l = MLP(LN(z_l)) + z_l \tag{3}$$

As shown in Fig. 5, in the transformer encoder layer, the updates are performed in two phases: a Multi-head Self-Attention (MSA) (Eq. (2)) and then a Multi-Layer Perceptron (MLP) (Eq. (3)). Both use residual skip connection and apply normalization to the sequence. In the MSA stage, the key elements of the transformer encoder, after getting the normalized input, the model starts by calculating the heads following

$$head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{4}$$

were $Q, KandV$ are matrices of $q_i = W^Q \cdot LN(z_{l-1}^i)$, $k_i = W^K \cdot LN(z_{l-1}^i)$ and $v_i = W^V \cdot LN(z_{l-1}^i)$ respectively, and W^Q, W^K and W^V are matrices to be learned.

$$MultiHead(Q, K, V) = Concat(head_1, \ldots, head_h)W^O \tag{5}$$

Then calculate the Multi-head function with W^O another matrix for the model to learn. After that, a two-layer MLP is applied. Note that the variants proposed in (Dosovitskiy et al., 2020) extend the number of the statistics parameters for "ViT-Large" and "ViT-Huge", hence the need for more computational power for the official variants.

3. Experiment setup

In this work, two main experiments were conducted. The first experiment was limited to the minimum number of images in the folder of a class. In our experiments, we set this number to 100 images per class since the minimum was 100 images in the "Damaged" class. Since we have six classes, the total number of images for the first experiment was 600 images to have the same number of images in each class, this to avoid having an unbalanced dataset. For the second experiment, we used all the data in an unbalanced dataset but in a stratified distribution to evaluate the dataset of 3921 images.

For the hyper-parameter search, a grid search was performed to select the best optimizer to use and the corresponding learning rate (LR). Each time, models were reset and a unique random combination of the parameters was used for the search. At the end of the test, only parameters that allow the model to achieve the best accuracy and lower loss compared to the previous results while searching were recorded and used for the experiments. The search uses 150 images from the dataset in a balanced configuration, and the training was limited to 5 epochs due to limited time and resources.

Before feeding the model, a resize to a shape of (224x224x3) was applied, with random horizontal flip and normalization. In addition, in all the experiments, we varied the test size by changing the number of

Table 2
Distribution of the 3921 manually labeled images.

Classes	Field 1	Field 2	Field 3	Field 4	Field 5	Field 6	Field 7	Field 8	Total
Damaged	13	0	2	27	16	7	30	5	**100**
EarlyStage	76	0	12	5	12	113	28	55	**301**
Healthy	21	0	0	9	3	47	18	25	**123**
LateStage	82	71	329	256	57	507	79	1244	**2625**
NoCactus	46	41	49	78	30	211	143	71	**669**
Old_Dead	11	0	10	20	3	31	3	25	**103**
Total	249	112	402	395	121	916	301	1425	**3921**

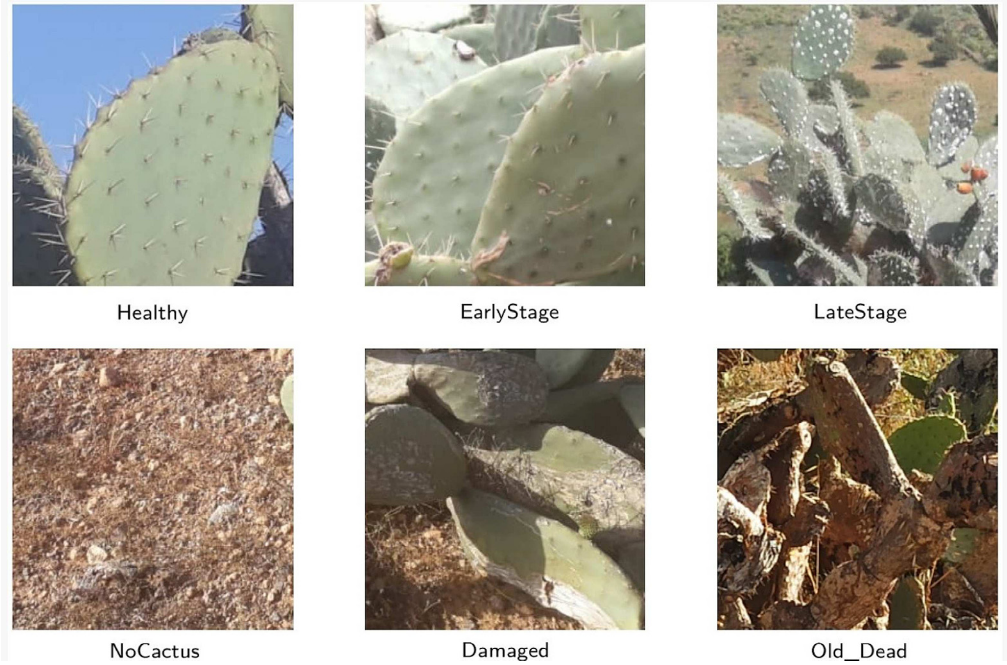


Fig. 4. Example of images from our dataset.

folds used for cross-validation of the overall datasets to validate the evaluations on all images. For example, by choosing 3 folds for training, the data were divided into 3 sets. These sets were selected in a stratified method using the Sklearn method "StratifiedKFold". Each is identical in the number of images per class without any data augmentation. One fold was used for the test (33%). The other two folds were combined and used for the training (67%) with the hyperparameters already selected. This training set was then split to search for the best set for validation using only ten epochs. Then we started the training process in 100 epochs using the best configuration for validation. When the model converged, we trained it again using the remaining validation set. We repeated the process for the following train-test sets, and calculated the overall performance of the results on the data.

Experiences were unified under the same condition for all tests using a local machine equipped with a Xeon(R) W-2123 @ 3.60GHz × 8 CPU, 32 Gb RAM, 500 Gb SSD and NVIDIA Corporation GP104GL [Quadro P4000] 8 Gb GDDR5 for training and evaluation of all experiences in the same global configuration. As for software, the Python 3 language was chosen because its libraries, such as Timm (Wightman, 2019), facilitate obtaining pre-trained models weights on public datasets to apply transfer learning.

The comparison between the models was evaluated using the log files of the evaluation of each model. Having the y_{pred} and y_{true} values, we were able to calculate the traditional metrics: accuracy, precision, recall, F1-score and the Matthews correlation coefficient to evaluate the confusion matrix (Zhu, 2020). All the metrics were used to obtain the model performance using the equations:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

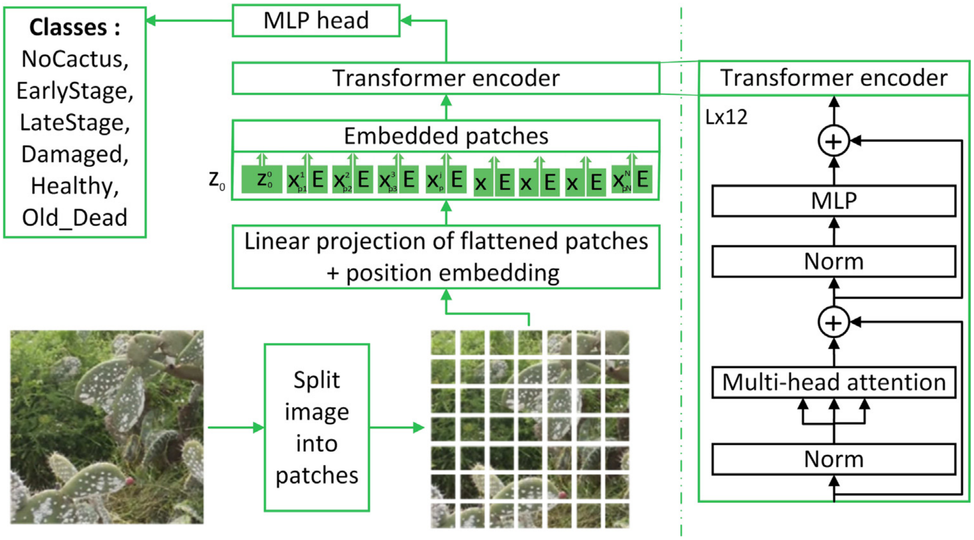


Fig. 5. The ViT-B architecture.

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{9}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \tag{10}$$

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

The setup we used for our CactiVIT mobile application was different. CactiVIT was built for Android 5.0 Lollipop (API 21) users to ensure that all users could use our app. The software application used to create it was Android Studio 2021.1.1 Patch 2 (Bumblebee). The coding was done using Python, Java, XML and PHP. The machine used for coding was equipped with an AMD Ryzen-76800H CPU @ 4.10GHz × 16, 16 Gb RAM, 500Gb SSD NVMe and NVIDIA RTX 3070Ti 8Gb GDDR6. For testing, we used virtual machines and an actual smartphone runnig android 11 equipped with a Mediatek Dimensity 700 2.2 GHz × 8 CPU, 6 Gb RAM and 128Gb ROM.

4. Results and discussion

4.1. Evaluation of ViT

Since ViT had multiple variations, in this subsection we focused on comparing them to each other in order to select the best ones to then compare with the rest of the CNN-based models. The criterion for choosing one variety over another was the performance obtained on our local machine. These tests revealed that due to resource limitations, the "ViT-Lage" and "ViT-Huge" architectures could not be evaluated. Using pre-trained weights, the models achieved high accuracy even though it was the first time they were run on this data.

After many experiences, the results in Fig. 6 show that the average score gave the advantage to the "ViT-B/16 mill" model variation. The score is based on the average of the metrics: Accuracy, Precision, Recall, F1-score and MCC. The selected model "ViT-B/16 mill" performed the best because it has the original "ViT-Base" architecture that uses 16 patches and was pre-trained by (Ridnik et al., 2021) on the ImageNet dataset: thus it performed better than the others due to the complexity of this dataset.

4.2. Comparison of ViT with other models

To compare the selected model "ViT-B/16 mill" with other CNN models, we performed our experiences on the local machine. The criterion was the same: models that could be tested on the local machine.

The results of the first experiment using transfer learning presented in Table 3 show the comparison of the selected models in various configurations. The ViT-B/16 model performed best using 50%, 33% and 25% of the dataset as test. But the ViT-B/32 outperformed it while using only 20% as test. This is because the ViT-B/32 had a larger input patches size in this configuration that required more data for training. Overall, for this balanced dataset of 600 images, the ViT-B/16 has an average accuracy of 73%, followed by the ResNet26 with 70.66% while the third was DenseNet169 with 70.25%.

The results of the second experiment presented in Table 4, show a higher performance due to the use of a larger dataset. We compared the same selected models in the various configuration, but this time using the large, unbalanced dataset of 3921 images. Once again, the ViT-B/16 model performed the best with an average accuracy of 88.73%. It was followed by the ViT-B/32_in21k model with 87.57% while the third is DenseNet169 with 87.04%.

We can assume from the results in Tables 3 and 4 that ViT performed best because it was trained on more data. Also, looking at the classification metrics for each class in Table 5 despite the low accuracy on half of the classes, the ViT had the best results we had seen. Conversely, CNN models like DenseNet can obtain acceptable results without transfer learning, but by adding this knowledge, additional performance was proven, but still inferior to the performance obtained by ViT. We can conclude that ViT, in general, achieves significant improvements after using transfer learning (TL) and a maximum amount of data for training.

4.3. Discussion

The experiences revealed in Tables 3 and 4 that having a larger dataset can improve the models' performances significantly. Interestingly, ViT remains in the lead in both experiences. The main difference is that ViT analyzes the entire image and then uses the attention

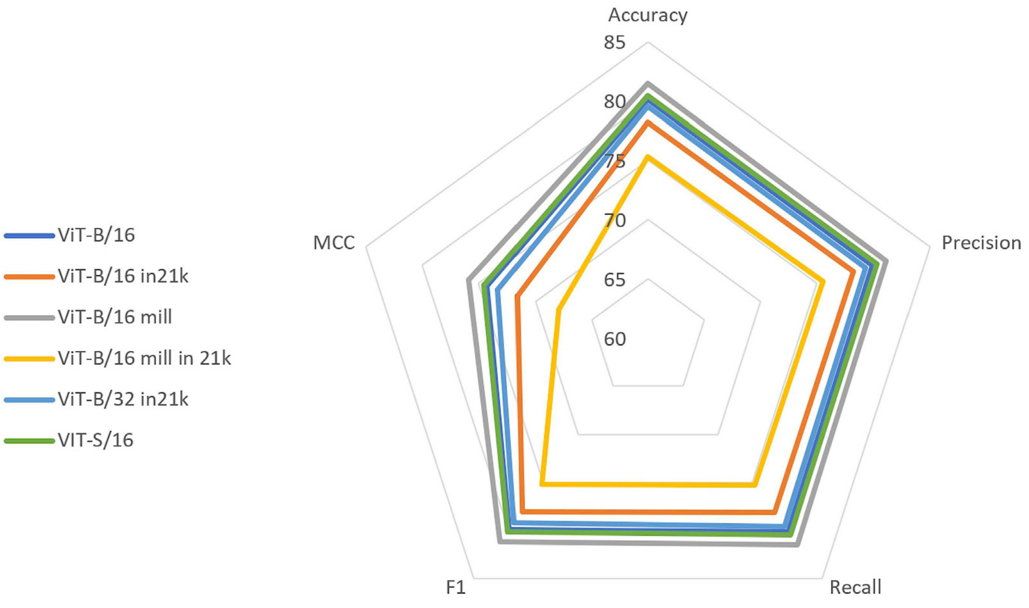


Fig. 6. Comparison of ViT variations.

Table 3
Results of the experimentation N°1 using our dataset of cacti limited to 100 images per class.

Experience N°1	20% Test		25% Test		33% Test		50% Test	
Model	Acc	F1	Acc	F1	Acc	F1	Acc	F1
DenseNet121	70.00	69.71	70.50	70.16	69.17	68.77	61.50	60.27
DenseNet161	73.17	72.82	68.83	68.76	70.50	70.23	65.67	64.40
DenseNet169	71.67	71.33	71.50	71.24	71.00	70.78	66.83	66.45
DenseNet201	72.00	71.84	68.50	67.79	69.83	69.59	66.33	66.00
EfficientNet_b0	68.67	68.30	68.67	68.15	65.83	65.49	58.50	58.38
EfficientNet_b1	69.67	69.45	63.67	62.89	65.33	65.22	60.00	58.36
EfficientNet_b2	67.83	67.94	66.67	66.51	67.17	67.21	57.50	57.22
Inception_Resnet_V2	69.67	69.51	63.17	63.28	67.50	67.13	61.00	61.02
Resnet18	71.50	71.35	70.17	69.82	70.17	69.83	66.17	65.88
Resnet26	71.17	70.76	73.00	72.61	70.17	69.95	68.33	67.92
Resnet34	69.50	69.29	69.33	69.00	66.83	66.25	61.67	60.87
ViT-B/16 mill	73.50	73.24	**74.00**	**73.66**	**74.00**	**73.80**	**70.50**	**70.11**
VIT-B32_in21k	**73.67**	**73.50**	70.50	70.46	72.50	72.31	63.67	62.79
Xception	72.67	72.37	70.00	69.50	69.67	69.47	65.33	64.99

Table 4
Results of the experimentation N°2 using all images in our dataset of cacti with unbalanced data.

Experience N°2	20% Test		25% Test		33% Test		50% Test	
Model	Acc	F1	Acc	F1	Acc	F1	Acc	F1
DenseNet121	87.57	87.53	87.19	87.04	86.81	86.61	85.00	84.48
DenseNet161	87.22	86.94	87.14	87.11	86.94	86.41	85.69	85.15
DenseNet169	87.68	87.62	87.75	87.45	87.32	86.92	85.43	84.86
DenseNet201	87.19	87.29	85.92	86.05	85.97	85.96	83.98	83.95
EfficientNet_b0	87.80	87.20	86.99	86.41	85.43	84.86	84.69	83.98
EfficientNet_b1	87.65	87.32	87.57	87.19	86.35	85.67	83.59	82.98
EfficientNet_b2	87.24	87.16	86.43	86.42	85.79	85.32	83.92	83.38
Inception_resnet_v2	83.26	83.37	84.41	84.13	84.05	83.48	82.39	82.13
Resnet18	87.78	87.37	86.45	86.17	86.55	86.08	84.56	83.94
Resnet26	86.94	87.10	86.12	86.01	85.97	85.62	84.49	84.18
Resnet34	86.32	86.26	85.40	84.95	85.33	84.85	83.01	82.49
ViT-B/16 mill	**88.62**	**88.45**	**89.26**	**88.78**	**88.70**	**88.29**	**88.34**	**87.42**
VIT-B32_in21k	88.01	87.45	88.47	87.81	87.85	87.20	85.94	84.62
Xception	87.55	87.40	87.65	87.49	86.35	86.28	80.10	80.70

mechanism to understand the semantics in the provided data, like natural language processing (NLP), which leads to better results. CNN, on the other hand, focuses locally on the neighbors of the pixels, so it may miss useful information. The problem arises when we have multi-features, such as having an early-stage cochineal infestation on the top of the cactus with a spiny rackets, in this cases the model cannot predict with a high accuracy. This precision is crucial for agents since we cannot wait for the cochineal to spread. The low accuracy mainly on Damaged, EarlyStage, and Healthy, is due to the similar features shared between them. While cochineal has the main characteristic of white wax on the leaves of cacti and is very bright in the late stage, in the early stage, indications that a cactus is infected can easily be mistaken for thorns with sun reflection or dust or even another infestation other than cochineal.

Table 5
ViT-B/16 mill performance.

Class	Experimentation N°1		Experimentation N°2	
	N° Images	Accuracy	N° Images	Accuracy
Damaged	100	55%	100	50%
EarlyStage	100	66%	301	73%
Healthy	100	57%	123	45%
LateStage	100	76%	2625	91%
NoCactus	100	99%	669	97%
Old_Dead	100	85%	103	53%

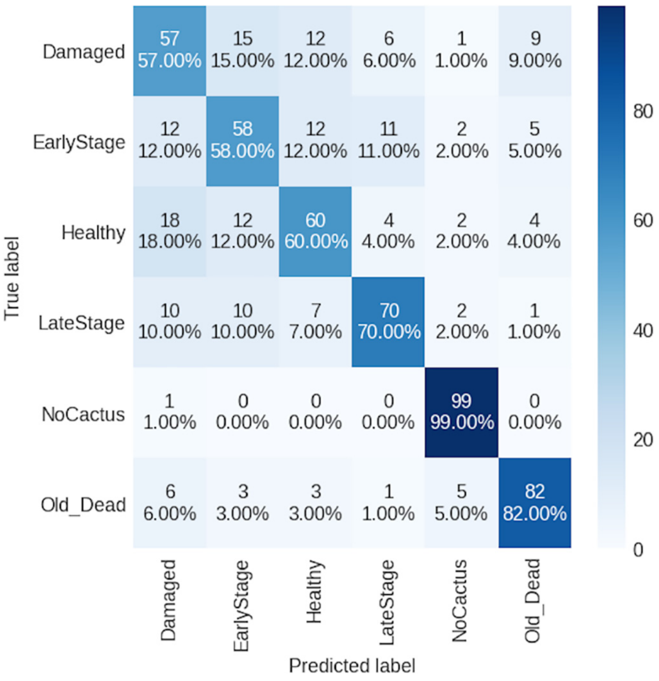


Fig. 7. Matrix confusion DensNet169.

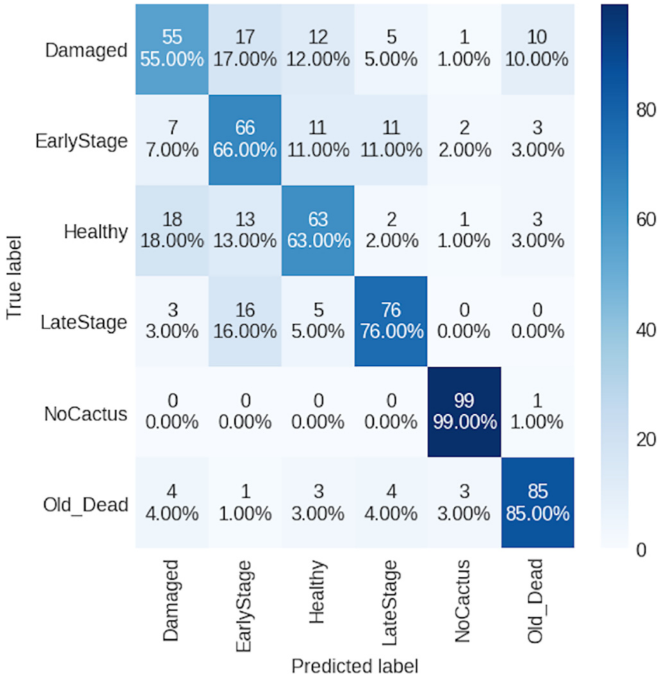


Fig. 8. Matrix confusion ViT.

In Figs. 7 and 8, we can observe that the main confusion is between the classes Damaged, EarlyStage and Healthy; this is due to the similarities between the images of these classes.

The main drawback of this data set, since it was acquired in the field, is the possibility of confusing the waxy symptoms of cochineal with thorns. The existence of these thorns can be found on both healthy and infected cacti (Fig. 9). In addition, images were not acquired under the same conditions, the image scale was random, and the angle of the shots was not fixed at constant values.

4.4. The mobile application: CactiViT

The proposed mobile application, "CactiViT," is specifically designed to assess the health status of cactus based on images captured by the user. It provides a more detailed classification by displaying the accuracy for every class for convenience use. Furthermore, the application allows users to provide valuable feedback on the classification results, facilitating ongoing improvements and refinements to the system.

The process of using the app is easy for farmers. After logging in, they just take a picture and obtain results. The main advantage of our mobile application is that it uses the serialized model of the original model. So, the farmer does not need to be connected to internet to obtain results. All processing is performed on the phone.

Fig. 10 illustrates CactiViT with its variant interfaces. Fig. 10.a shows the main graphical user interface (GUI) with the application's home page: the user can sign in, log in or just use the application as a guest.

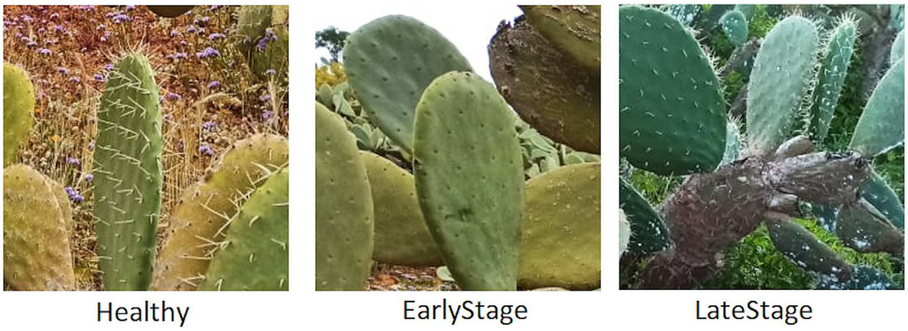
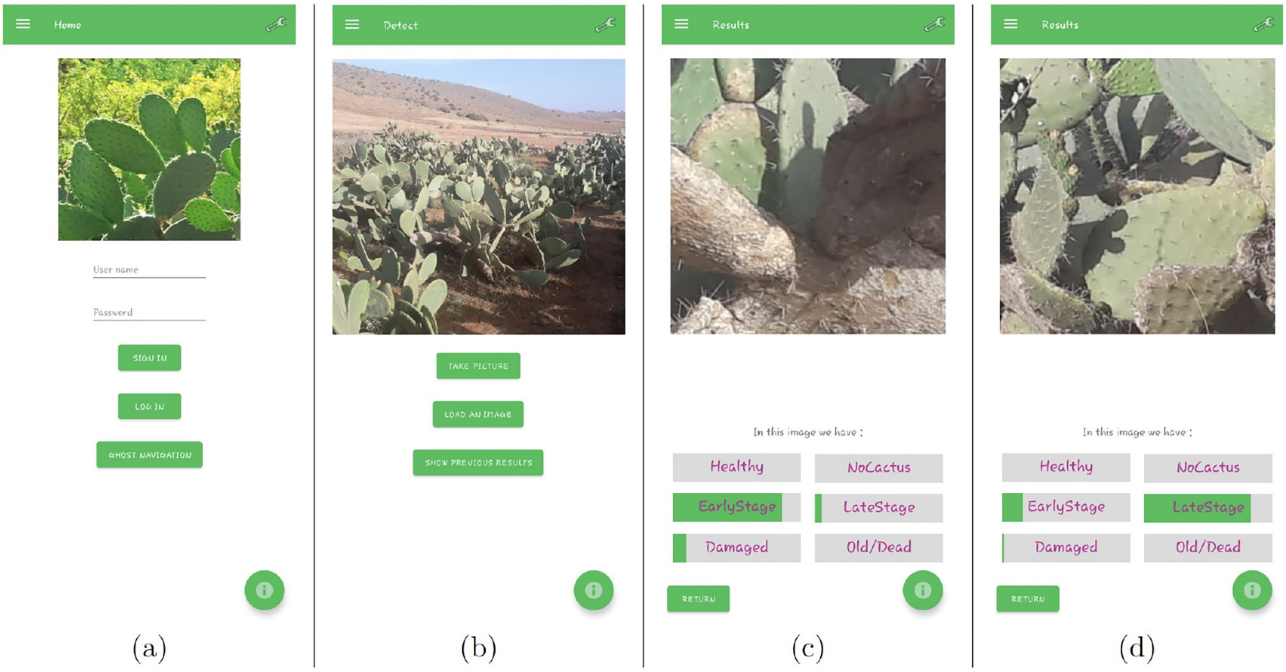


Fig. 9. Possibilities of confusion.



In this image we have :

In this image we have :

Fig. 10. Our CactiVIT GUI.

In Fig. 10.b, the user can either take a new picture, load an image from the gallery or simply review the previously obtained results to give a feedback. CactiViT, provides an insight into the classes probabilities associated with each health status of an image. The example on Fig. 10.c CactiViT indicates that the image corresponds to a cactus in the early stage of infection. While in Fig. 10.d CaciViT predicts a late-stage infection.

5. Conclusion

Diagnosing the health statuses for vegetation product using only mobile phone images is the main goal of preventing losses and improving the quality and quantity of plant resources. The proposed application can help farmers to have an idea about the cactus in remote field by showing the probabilities of the health statuses, in case of an infestation exist the farmer can take decision for an early intervention. It has a complete system that acquires data, train it on a remote machine using the ViT model, notifies landowners of a developing infestation and helps conserve resources. In this paper, we have contributed with the application and the dataset used so that researchers can make further improvements. In addition, we have compared the classification models on our dataset, and ViT-B/16 was the best performing model with a score of 88.73%, followed by ViT-B/32_in21k with 87.57%; in comparison, the DenseNet169 model scored 87.04%. In comparison, the Densnet169 scored 87.04%. To further our research, we plan to create a better performing model with low environmental impact. Further studies will focus on serialization to deploy a better mobile application. In addition, it will be generalized to monitor other crops.

CRediT authorship contribution statement

Anas Berka: Software, Writing – original draft, Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Investigation. **Adel Hafiane:** Supervision, Writing – review & editing, Conceptualization, Methodology, Validation. **Youssef Es-Saady:** Supervision, Writing – review & editing, Conceptualization, Methodology, Validation, Investigation. **Mohamed El Hajji:** Supervision, Writing – review & editing, Conceptualization, Methodology, Validation, Investigation. **Raphaël Canals:** Supervision, Writing – review & editing, Conceptualization, Methodology, Validation. **Rachid Bouharroud:** Writing – review & editing, Validation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was made possible by project PHC Toubkal/21/121 – Campus France: 45942UG who funded the work and the mobility fee between Morocco and France. We also thank Mr. Marouane EL KAOURI for his ongoing collaboration to create the image database.

References

Akroud, H., Sbaghi, M., Bouharroud, R., Koussa, T., Boujghagh, M., El Bouhssini, M., 2021. Antibioisis and antixenosis resistance to dactylopius opuntiae (hemiptera: Dactylopiidae) in moroccan cactus germplasm. Phytoparasitica 623–631. https://doi.org/10.1007/s12600-021-00897-w.

Aldama, A.C., Celina, L., Soto Hernandez, M., Castillo-Márquez, L., 2005. Cochineal (dactylopius coccus costa) production in prickly pear plants in the open and in microtunnel greenhouses. Agrociencia 39, 161–171.

Amani, E., Marwa, L., Hichem, B.S., Amel, S.H., Ghada, B., 2019. Morphological variability of prickly pear cultivars (opuntia spp.) established in ex-situ collection in Tunisia. Sci. Hortic. (Amsterdam) 248, 163–175.

Andrianto, H., Suhardi Faizal, A., Armandika, F., 2020. Smartphone application for deep learning-based rice plant disease detection. 2020 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE.

Atila, U., Uçar, M., Akyol, K., Uçar, E., 2021. Plant leaf disease classification using EfficientNet deep learning model. Ecol. Inform. 61, 101182. https://doi.org/10.1016/j.ecoinf.2020.101182.

Atitallah, S.B., Driss, M., Boulila, W., Koubaa, A., Atitallah, N., Ghézala, H.B., 2021. An enhanced randomly initialized convolutional neural network for columnar cactus recognition in unmanned aerial vehicle imagery. Proc. Comp. Sci. 192, 573–581. https://doi.org/10.1016/j.procs.2021.08.059.

Barman, U., Choudhury, R.D., Sahu, D., Barman, G.G., 2020. Comparison of convolution neural networks for smartphone image based real time classification of citrus leaf disease. Comput. Electron. Agric. 177, 105661.

Bouharroud, R., Amarraque, A., Qessaoui, R., 2016. First report of the opuntia cochineal scale dactylopius opuntiae (hemiptera: Dactylopiidae) in Morocco. Bull. OEPP 46, 308–310.

Brahimi, M., Boukhalfa, K., Moussaoui, A., 2017. Deep learning for tomato diseases: classification and symptoms visualization. Appl. Artif. Intell. 31, 299–315. https://doi.org/10.1080/08839514.2017.1315516.

Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., Moussaoui, A., 2018. Deep learning for plant diseases: Detection and saliency map visualisation. Human and Machine Learning. Springer International Publishing, pp. 93–117.

Carton, B., Mongardini, M.J., Li, Y., 2018. A New Smartphone for every Fifth Person on Earth: Quantifying the New Tech Cycle. International Monetary Fund.

DiMiceli, C., Townshend, J., Carroll, M., Sohlberg, R., 2021. Evolution of the representation of global vegetation by vegetation continuous fields. Remote Sens. Environ. 254, 112271. https://doi.org/10.1016/j.rse.2020.112271.

Dodd, A.P., et al., 1940. The Biological Campaign against Prickly-Pear. Commonwealth Prickly Pear Board, Brisbane.

Donkin, R.A., 1977. Spanish red: an ethnogeographical study of cochineal and the opuntia cactus. Trans. Am. Philos. Soc. 67, 1. https://doi.org/10.2307/1006195.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. http://arxiv.org/abs/2010.11929.

Durmuş, H., Güneş, E.O., Kırcı, M., 2017. Disease detection on the leaves of the tomato plants by using deep learning. 2017 6th International Conference on Agro-geoinformatics. IEEE, pp. 1–5.

El Massi, I., Es-saady, Y., El Yassa, M., Mammass, D., 2021. Combination of multiple classifiers for automatic recognition of diseases and damages on plant leaves. Sign. Image Video Proc. https://doi.org/10.1007/s11760-020-01797-y.

Fang, Y., Ramasamy, R.P., 2015. Current and prospective methods for plant disease detection. Biosensors 5, 537–561. https://doi.org/10.3390/bios5030537.

Ferentinos, K.P., 2018. Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. 145, 311–318. https://doi.org/10.1016/j.compag.2018.01.009.

Foldi, I., 2001. Liste des cochenilles de france (hemiptera, coccoidea). Bulletin de la Societe entomologique de France 106, 303–308. https://doi.org/10.3406/bsef.2001.16768.

Foxcroft, L.C., Hoffmann, J.H., 2000. Dispersal of *Dactylopius opuntiae* cockerell (homoptera: Dactylopiidae), a biological control agent of *Opuntia stricta* (haworth.) haworth. (cactaceae) in the kruger national park. Koedoe 43.

Gao, X., Qian, Y., Gao, A., 2021. COVID-VIT: classification of COVID-19 from CT chest images based on vision transformer models. arXiv https://doi.org/10.48550/ARXIV.2107.01682 preprint.

Gheflati, B., Rivaz, H., 2022. Vision transformers for classification of breast ultrasound images. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE, pp. 480–483 https://doi.org/10.1109/EMBC48229.2022.9871809.

GmbH, P., 2019. Plantix. The Smart Crop Assistant on Your Smartphone. https://plantix.net/en/.

Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Yahiaoui, I., Selmi, S., Carré, J., Barthélémy, D., Boujemaa, N., Molino, J.F., Duché, G., Péronnet, A., 2013. Pl@ntNet Mobile App. , pp. 423–424 https://doi.org/10.1145/2502081.2502251.

Golhani, K., Balasundram, S.K., Vadamalai, G., Pradhan, B., 2018. A review of neural networks in plant disease detection using hyperspectral data. Inform. Proc. Agric. 5, 354–371. https://doi.org/10.1016/j.inpa.2018.05.002.

Griffith, M.P., 2004. The origins of an important cactus crop, opuntia ficus-indica (cactaceae): new molecular evidence. Am. J. Bot. 91, 1915–1921.

Hughes, D., Salathé, M., et al., 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint http://arxiv.org/abs/1511.08060.

Icrisat, 2019. Mobile App to Help Farmers Overcome Crop Damage. http://www.icrisat.org/mobile-app-to-help-farmers-overcome-crop-damage/.

Kaur, B., Sharma, T., Goyal, B., Dogra, A., 2020. A genetic algorithm based feature optimization method for citrus HLB disease detection using machine learning. 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE, pp. 1052–1057.

Kaweesinsakul, K., Nuchitprasitchai, S., Pearce, J., 2021. Open source disease analysis system of cactus by artificial intelligence and image processing. The 12th International Conference on Advances in Information Technology, pp. 1–7.

Kerkech, M., Hafiane, A., Canals, R., 2020. VddNet: Vine Disease Detection Network Based on Multispectral Images and Depth Map. http://arxiv.org/abs/2009.01708.

Krizhevsky, A., Sutskever, I., Hinto, G.E., 2012. 2012 AlexNet. Adv. Neural Inf. Proces. Syst. 281, 1–9.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90. https://doi.org/10.1145/3065386.

Le Houérou, H.N., 1996. The role of cacti (opuntiaspp.) in erosion control, land reclamation, rehabilitation and agricultural development in the mediterranean basin. J. Arid Environ. 33, 135–159. https://doi.org/10.1006/jare.1996.0053.

Li, Z., Wallace, E., Shen, S., Lin, K., Keutzer, K., Klein, D., Gonzalez, J.E., 2020. Train Large, then Compress: Rethinking Model Size for Efficient Training and Inference of Transformers. http://arxiv.org/abs/2002.11794.

Liang, J., Wang, D., Ling, X., 2021. Image classification for soybean and weeds based on ViT. J. Phys. Conf. Ser. 2002, 012068. https://doi.org/10.1088/1742-6596/2002/1/012068.

Lotto, G.D., 1974. On the status and identity of the cochineal insects (homoptera: Coccoidea: Dactylopiidae). J. Entomol. Soc. South. Afr. 37, 167–193.

MacDonald, S.L., Staid, M., Staid, M., Cooper, M.L., 2016. Remote hyperspectral imaging of grapevine leafroll-associated virus 3 in cabernet sauvignon vineyards. Comput. Electron. Agric. 130, 109–117.

Machado, B.B., Orue, J.P., Arruda, M.S., Santos, C.V., Sarath, D.S., Goncalves, W.N., Silva, G.G., Pistori, H., Roel, A.R., Rodrigues-Jr, J.F., 2016. BioLeaf: a professional mobile application to measure foliar damage caused by insect herbivory. Comput. Electron. Agric. 129, 44–55.

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., Davis, C.E., Dandekar, A.M., 2015. Advanced methods of plant disease detection. A review. Agron. Sustain. Dev. 35, 1–25. https://doi.org/10.1007/s13593-014-0246-1.

Miller, D.R., 1996. Checklist of the scale insects (coccoidea: Homoptera) of mexico. Proc. Entomol. Soc. Wash. 98, 68–86.. https://www.biodiversitylibrary.org/part/66608.

Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7, 1419. https://doi.org/10.3389/fpls.2016.01419.

Ngugi, L.C., Abelwahab, M., Abo-Zahhad, M., 2021. Recent advances in image processing techniques for automated leaf pest and disease recognition – a review. Inform. Proc. Agric. 8, 27–51. https://doi.org/10.1016/j.inpa.2020.04.004.

Nobel, P.S., 2002. Cacti: Biology and Uses. University of California Press.

Ouhami, M., Es-Saady, Y., Hajji, M.E., Hafiane, A., Canals, R., Yassa, M.E., 2020. Deep transfer learning models for tomato disease detection. Lecture Notes in Computer Science. Springer International Publishing, pp. 65–73.

Ouhami, M., Hafiane, A., Es-Saady, Y., El Hajji, M., Canals, R., 2021. Computer vision, IoT and data fusion for crop disease detection using machine learning: a survey and ongoing research. Remote Sens. 13. https://doi.org/10.3390/rs13132486.

Perez, M.F., Bonatelli, I.A., Romeiro-Brito, M., Franco, F.F., Taylor, N.P., Zappi, D.C., Moraes, E.M., 2022. Coalescent-Based Species Delimitation Meets Deep Learning: Insights from a Highly Fragmented cactus System. Molecular Ecology Resources 22, 1016–1028. Publisher: Wiley Online Library.

Petrellis, N., 2017. A smart phone image processing application for plant disease diagnosis. 2017 6th International Conference on Modern Circuits and Systems Technologies (MOCAST). 1–4. IEEE.

Pongnumkul, S., Chaovalit, P., Surasvadi, N., 2015. Applications of smartphone-based sensors in agriculture: a systematic review of research. J. Sens. 2015, 1–18. https://doi.org/10.1155/2015/195308.

Qiang, Z., He, L., Dai, F., 2019. Identification of plant leaf diseases based on inception v3 transfer learning and fine-tuning. Communications in Computer and Information Science. Springer Singapore, pp. 118–127 (ISSN: 1865-0929).

Ramesh, S., Hebbar, R., Niveditha, Pooja, Bhat, P., Shashank, Vinod, 2018. Plant disease detection using machine learning. In: 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C). IEEE.

Reddy, S., Pawar, A., Rasane, S., Kadam, S., 2015. A survey on crop disease detection and prevention using android application. Int. J. Innov. Sci. Eng. Technol. 2, 621–626.

Reedha, R., Dericquebourg, E., Canals, R., Hafiane, A., 2022. Transformer neural network for weed and crop classification of high resolution UAV images. Remote Sens. 14, 592. https://doi.org/10.3390/rs14030592.

Ridnik, T., Ben-Baruch, E., Noy, A., Zelnik-Manor, L., 2021. ImageNet-21k Pretraining for the Masses.

Shrivastava, V.K., Pradhan, M.K., Minz, S., Thakur, M.P., 2019. Rice plant disease classification using transfer learning of deep convolution neural network. Int. Archiv. Photogram. Rem. Sens. Spat. Inform. Sci. 3, 631–635.

Spodek, M., Ben-Dov, Y., Protasov, A., Carvalho, C.J., Mendel, Z., 2014. First record of dactylopius opuntiae (cockerell)(hemiptera: Coccoidea: Dactylopiidae) from Israel. Phytoparasitica 42, 377–379.

Stintzing, F.C., Carle, R., 2005. Cactus stems (opuntia spp.): a review on their chemistry, technology, and uses. Mol. Nutr. Food Res. 49, 175–194. https://doi.org/10.1002/mnfr.200400071.

Thai, H.T., Tran-Van, N.Y., Le, K.H., 2021. Artificial cognition for early leaf disease detection using vision transformers. 2021 International Conference on Advanced Technologies for Communications (ATC), pp. 33–38 https://doi.org/10.1109/ATC52653.2021.9598303.

Tigano, A., Colella, J.P., MacManes, M.D., 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Mol. Ecol. 29, 1300–1314.

Too, E.C., Yujian, L., Njuki, S., Yingchun, L., 2019. A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture 161, 272–279. Publisher: Elsevier.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., Jégou, H., 2020. Training Data-efficient Image Transformers & Distillation Through Attention. http://arxiv.org/abs/2012.12877.

Vasconcelos, A.G.V.D., Lira, M.D.A., Cavalcanti, V.L.B., Santos, M.V.F.D., Willadino, L., 2009. Seleção de clones de palma forrageira resistentes Ã cochonilha-do-carmim (dactylopius sp). Rev. Bras. Zootec. 38, 827–831. https://doi.org/10.1590/S1516-35982009000500007.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is All You Need. http://arxiv.org/abs/1706.03762.

Wang, A., Zhang, W., Wei, X., 2019. A review on weed detection using ground-based machine vision and image processing techniques. Comput. Electron. Agric. 158, 226–240. https://doi.org/10.1016/j.compag.2019.02.005.

Wightman, R., 2019. PyTorch Image Models. https://doi.org/10.5281/zenodo.4414861.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: State-of-the-art natural language processing. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Association for Computational Linguistics, pp. 38–45.

Xie, C., Yang, C., He, Y., 2017. Hyperspectral imaging for classification of healthy and gray mold diseased tomato leaves with different infection severities. Comput. Electron. Agric. 135, 154–162.

Yuan, Y., Chen, L., Wu, H., Li, L., 2022. Advanced agricultural disease image recognition technologies: a review. Inform. Proc. Agric. 9, 48–59. https://doi.org/10.1016/j.inpa.2021.01.003.

Zhai, X., Kolesnikov, A., Houlsby, N., Beyer, L., 2021. Scaling Vision Transformers. http://arxiv.org/abs/2106.04560.

Zhang, K., Zhao, J., Zhu, Y., 2018. MPC case study on a selective catalytic reduction in a power plant. J. Process Control 62, 1–10. https://doi.org/10.1016/j.jprocont.2017.11.010.

Zhu, Q., 2020. On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset. Pattern Recogn. Lett. 136, 71–80. https://doi.org/10.1016/j.patrec.2020.03.030.

Zhu, M., Tang, Y., Han, K., 2021. Vision Transformer Pruning. http://arxiv.org/abs/2104.08500.