Computers and Electronics in Agriculture 226 (2024) 109478

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

In-field disease symptom detection and localisation using explainable deep learning: Use case for downy mildew in grapevine

Inés Hernández [a,b], Salvador Gutiérrez [c], Ignacio Barrio [a,b], Rubén Íñiguez [a,b], Javier Tardaguila [a,b,*]

[a] Televitis Research Group, University of La Rioja, 26006 Logroño, Spain
[b] Institute of Grapevine and Wine Sciences (University of La Rioja, Consejo Superior de Investigaciones Científicas, Gobierno de La Rioja), 26007 Logroño, Spain
[c] Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain

ARTICLE INFO

Keywords:
Disease detection
Computer vision
Convolutional neural networks
Vision transformers
Explainable artificial intelligence
Digital agriculture

ABSTRACT

Diseases and pests in agriculture significantly impact crop yield and quality. Downy mildew (*Plasmopara viticola*) is a particular noteworthy example in grapevines. Traditional detection methods are laborious, subjective and time-consuming. Consequently, a technological solution based on artificial intelligence, would provide higher levels of reproducibility and sampling. The aim of this work was to develop an interpretable, automated method for detection and localisation of plant disease symptoms under field conditions. Images of the grapevine canopy were taken in 14 commercial vineyard plots under a range of lightning conditions, including both static and on-the-go settings. The images were processed using a sliding window, classifying sub-images into areas with and without downy mildew. Transfer learning, fine-tuning and data augmentation were employed to automate the classification, comparing convolutional neural networks (CNNs) and vision transformers (ViT). Subsequently, the trained model was integrated into the sliding window to localise regions within the canopy images exhibiting symptoms of downy mildew. Model predictions were interpreted using explainable artificial intelligence (XAI) methods. The EfficientNetV2S model achieved an accuracy of 91 % and an F1-score of 0.92 when classifying image areas and an Intersection over Union (IoU) of 0.83 when locating symptomatic areas. This method showed promising results, enabling automatic and explainable detection and localisation of plant diseases in complex conditions. The straightforward labelling process facilitated adaptation to new conditions, making it suitable for different crops and diseases. Integration into mobile platforms could enhance disease management and reduce the spread of pathogens, making a significant advance in agricultural technology.

1. Introduction

Diseases and pests are one of major problems in agriculture, whose effect could range from slowing plant growth to the death of the whole crop, directly affecting the quality and quantity of its production (Liu and Wang, 2021). Crop monitoring is relevant to prevent the spread of diseases and pests, reducing crop damage and the need for pesticides and other control measures. Traditional methods for detecting diseases and pests rely on visual inspection by trained experts, which can be time-consuming, subjective, and susceptible to human error (Paulus et al., 1997). Downy mildew (*Plasmopara viticola*) is one of the most devastating diseases affecting grapevine worldwide (Wilcox et al., 2015). This disease can present symptoms in the fruit, branch, and leaves of the plants. Detecting the symptoms in leaves is crucial because it is one of the primary sites of the infection and can be an early indication of the presence of the disease. The small size of the initial symptoms, which include oil spots and sporulation primarily on the leaf undersides, makes them difficult to identify (Gessler et al., 2011). Early identification in the field enables timely fungicide applications, thereby maximizing efficacy and preventing extensive crop loss. Hence, one of the main challenges in plant protection is automated disease symptoms detection under field conditions (Lee and Tardaguila, 2023; Liu et al., 2022; Zhang et al., 2022).

Non-invasive technologies are frequently used in phytopathology and crop protection, avoiding to damage crops during monitoring. Spectral signals (Al-Saddik et al., 2018; Nguyen et al., 2021) and thermal images (Mastrodimos et al., 2019) have been analysed to detect infections in crops such as grapevine. However, RGB image analysis

* Corresponding author.
E-mail address: javier.tardaguila@unirioja.es (J. Tardaguila).

https://doi.org/10.1016/j.compag.2024.109478
Received 27 May 2024; Received in revised form 18 September 2024; Accepted 20 September 2024
Available online 26 September 2024

remains one of the main techniques used for visual disease detection, due to its acquisition and processing simplicity and low economical cost. Similarly to the symptoms of other diseases such as potato late blight (Gao et al., 2021), or rice (Dey et al., 2022), mango (Kusrini et al., 2020), soybean (Bevers et al., 2022) or apple leaf diseases (Li et al., 2021), the symptoms of downy mildew on grapevine leaves (appearing as oil spots) represent a small part of the plant, so their visual localisation in early stages is complicated. For this reason, RGB images of leaves or plant regions are often used to detect diseases automatically with computer vision techniques or machine learning models. Disease symptoms can be detected in the laboratory or using simple backgrounds analysing image features using computer vision (Barbedo, 2014; Hernandez et al., 2022, 2021), but the use of more advanced techniques such as deep learning makes it possible to detect or differentiate them from other diseases or pests under more complex conditions, such as their analysis directly in the field (Bevers et al., 2022; Cai et al., 2023; Dey et al., 2022; Gutierrez et al., 2021). However, creating a method using self-collected images taken in the field remains a challenge, due to the changes in lighting that are not contemplated in the methods, background noise that might interfere with other objects of interest, or the small size of the disease lesions, often ignored in low-resolution images. It is remarkable that, although deep learning models require a lot of data, transfer learning and fine-tuning combined with data augmentation techniques are frequently used for complex and small datasets (Kusrini et al., 2020; Thakur et al., 2022). convolutional neural networks (CNNs) and vision transformers (ViTs) have been recognized as powerful tools for image analysis in recent years, being used in tasks such as the classification of leaf diseases in cassava (Thai et al., 2021), apple (Li et al., 2021), tomato (Agarwal et al., 2020) or rice (Zhou et al., 2023). CNNs have been the most widely used architecture for image classification due to their ability to automatically learn and extract features from images. However, while CNNs may have limitations in capturing global information in an image, ViTs have shown great potential in overcoming these limitations by processing the entire image as a sequence of patches, allowing them to access global information at once. Furthermore, explainable artificial intelligence (XAI) methods such as Gradient-weighted Class Activation Mapping (Grad-CAM) and attention maps have gained importance in recent years as they enable researchers to understand the reasoning behind the classification results of deep learning models (Barredo Arrieta et al., 2020; Kakogeorgiou and Karantzalos, 2021). The combination of these techniques has the potential to enhance the accuracy and interpretability of image classification models, making them more useful in various domains, including agriculture and medicine (Pintelas et al., 2021; Xia et al., 2021).

On the other hand, lesion localisation usually requires laborious annotation by experts which makes these methods more complex. For instance, Tardif et al. (2022) reported that it can take up to 20 min to locate different leaf diseases per grapevine image using bounding boxes. Ahmad et al., (2023) also identified the labour-intensive nature of accurately delineating the shape of the symptoms in image segmentation tasks. To address this challenge, the methodology presented in this work proposes a sliding window method that uses a deep neural network as feature extractor and classifier to achieve an easily adaptable method that allows extracting the location of symptomatic plant areas. Furthermore, this localisation can be interpreted using XAI methods, which avoids interpretability problems usually associated with this type of network, as is mentioned by Liu and Wang (2021). The use of a simple and intuitive method to locate symptomatic lesions, such as the sliding window, could provide a complete picture of the infection in the plant employing a simple annotation, rather than just a part of it, as most of the works in this field reflect (Liu and Wang, 2021).

This study introduces an innovative technology designed for the detection and localisation of foliar diseases and pests in plants under real-world conditions, accompanied by novel tools for the analysis of visual symptom detection. In particular, the methodology was demonstrated through the example of downy mildew detection in grapevines. The study examined the effectiveness of the benchmark CNNs and vision transformer in determining the optimal classification neural network for the detection of disease-affected regions in plants. To enhance confidence and usability, explainable artificial intelligence techniques were employed to interpret the predictions of the neural networks. In addition, the classification model was used to identify the regions of the plant affected by the disease through the use of a sliding window approach, thereby facilitating the localisation and quantification of the disease across the entire plant with a straightforward labelling methodology. Finally, the study analysed the adaptability of the method to varying daylight conditions, plant varieties and image acquisition, enabling its implementation on a mobile platform for accurate automated disease diagnosis and treatment in commercial crops.

2. Related works

Several studies have explored the use of deep learning to detect plant diseases and pests in real natural environments, with a focus on preventing and controlling their spread and development (Liu and Wang, 2021; Thakur et al., 2022). Deep neural networks are mainly used for feature extraction, classification and lesion localisation using data collected in the laboratory and in the field (Liu and Wang, 2021). In addition, image datasets are usually collected manually, using a simple background and focusing the image on the region of interest of the plant (Araujo and Peixoto, 2019; Chen et al., 2021), or the background may be even removed to decrease image complexity (Dey et al., 2022). which leads to simpler image processing and accurate results, but requires additional time.

Kumar Sahu and Pandey (2023) employ computer vision and machine learning techniques to differentiate plant foliar diseases in different crops. Image processing and Fuzzy C-Means clustering are used to extract features from the images and Random Forest (RF) and Multi-Class Support Vector Machine (MCSVM) models are hybridised to classify the data. The method used is well adapted to images taken under controlled conditions using a simple annotation of the images and allows an interpretation of the prediction due to the simplicity of the methods applied. However, its adaptation to complex field conditions would require more complex methods, such as deep learning. Arumuga Arun and Umamaheswari (2023) propose a Complete Concatenated Deep Learning (CCDL) architecture for crop disease classification, using a resampled version of the Plant Village dataset to pre-train a model capable of adapting to crop field images retraining the model through fine-tuning. The importance of model interpretability is also highlighted by visualizing the feature maps generated by the different levels of convolution layers. Although this model is notable for its potential for adaptation to resource-constrained devices, models that achieve a higher accuracy such as the Inception architecture, which could be relevant for disease detection, are discarded. In addition, the method focuses on images of specific parts of the leaves. These methods could be improved for efficient application in the field by considering image acquisition using a mobile platform, in different daylight conditions and focusing on the whole plant.

In the case of grapevine, techniques such as computer vision, deep neural networks, transfer learning, fine-tuning and data augmentation could be applied to accurately detect the presence of leaf disease symptoms. Wu et al. (2020) develop a computer vision method that extracts lesion pixels using spatial colour transformations, Wiener filtering, the Otsu method, morphological operations, and Prewitt edge detector, and then extracts various features from the size and shape of the lesions for their classification using a back-propagation neural network. This type of feature extraction is usually specialised in the datasets and requires intensive supervision by experts to adjust the parameters of each technique and is more useful in controlled conditions, as considered in the study. Other works consider deep learning to deal with natural environmental conditions, although in some cases computer vision techniques such as the use of HSV colour space and

morphology transformations are employed to simplify the region of interest, although deep learning is used to extract the lesion features (Gutierrez et al., 2021; Liu et al., 2022). However, the use of these techniques could be replaced by deep neural networks to detect leaf diseases such as downy mildew, as in the case of Zhang et al. (2022), who combine YOLOv5 with attention mechanisms. Meanwhile, Chen and Wu (2023) identify grape leaf diseases by augmenting their dataset using a Faster R-CNN and a generative adversarial network (DCGAN) to generate synthetic grape lesions, and then classify the images with a CNN. Although the findings appear promising, these approaches are limited to identifying diseases on leaves without considering the complexity of detecting diseases directly on the plants.

3. Materials and methods

Before exposing the specific details of this study, a previous, general overview is presented to the reader in Fig. 1. First, plant images were acquired in the field using portable digital cameras (a). Afterwards, raw data processing was performed involving image supervision, splitting, and adequately labelling by human experts (b). The next step was the deep learning modelling with neural networks for automated disease detection in leaves, trained with the images previously labelled (c). Modelling was supported and evaluated with XAI methods. Finally, affected leaves were located generating a sliding-window detector, capable of batch-processing canopy images for disease localisation (d).

3.1. Data acquisition

In-field RGB image acquisition was performed in fourteen commercial vineyard plots trained on vertical shoot positioning located in two regions in northern Spain presenting different levels of downy mildew incidence (Table 1). Image acquisition was carried out under different sky light conditions in two seasons (2019 and 2021).

Images were taken both manually and on-the-go at a height of approximately one metre from the ground and at 1.0 m from the vineyard canopy, ensuring the acquisition of the main body of the canopy. A ground mobile platform with permanent all-wheel drive (AS 940 Sherpa 4WD XL, AS-Motor GmbH, Bühlertann, Germany) was used to take the images on-the-go (Fig. 2). The vehicle was taking images at a speed of 5 Km/h, using a Canon EOS 5D Mark IV RGB camera (Canon Inc. Tokyo, Japan), mounting a full-frame CMOS sensor (35 mm and 30.4MP equipped with a Canon EF 20 mm F/2.8 USM lens). The images taken manually were obtained both with the Canon camera and with a Sony alpha 7-II digital mirrorless RGB camera (Sony Corp., Tokyo, Japan) equipped with a Zeiss 24/70 mm lens with optical stabilization. Images taken with the Canon camera had dimension of 6720 × 4480 pixels, while the ones taken with the Sony were 6000 × 4000 pixels. In total, 224 images were captured from all the plots: 52 on-the-go, 66 manually using the Canon camera, and 106 manually using the Sony camera (Table 1). Examples of the images taken with different natural lighting and shading can be seen in the Fig. 3.

3.2. Data preparation and labelling

Visual symptoms of downy mildew are mainly represented by small spots on leaves, which requires high-quality images for detection. Canopy of the plant was captured almost completely and similarly in all plots, due to its vertical shoot positioning system, which allowed the images to be processed in the same way, detecting which parts of the plants were diseased using a sliding window. The height and width of the leaves in the images varied from 1200 to 120 pixels, due to the variability in grapevine, the variation in both leaf size and leaf surface (which means that some shoots are closer to the camera than others or that there are variations in the age of the leaves) on each plant. For this reason, the dataset used to detect the downy mildew symptoms was created extracting sub-images of 800x800 pixels with a sliding window, considering approximately one leaf per window or an area with several leaves. The sliding window used to extract the data applied a 400x400 pixel stride (50 % overlap) to obtain more windows per image and to increase available data. For each 6720x4480 pixel image, 176 windows were extracted and for 6000x4000 pixel images, 126 windows were extracted. The extracted sub-images were resized to 320x320 pixels in order to adapt them to the input of the neural network models. This process reduced the computational load of the models while maintaining the texture of the disease symptoms in the images, despite the loss of detail.

The windows were manually classified by an expert into two classes (areas with downy mildew symptoms and areas without downy mildew symptoms) (Fig. 4). The infected areas included at least one leaf containing oil spots caused by downy mildew. The non-infected areas included all other cases, such as leaves without symptoms of downy mildew, the trunk of the plant, the soil, or the sky.

3.3. Disease detection

The sub-images were classified using six deep learning models and transfer learning, reducing training time transferring weights learned from a huge image dataset. The tested architectures were five: four CNNs and a vision transformer, all of them pre-trained using the ImageNet dataset (Russakovsky et al., 2015) and changing the original last dense layer to one with two neurons and a softmax activation function, to adapt the network to a binary classification problem. The four CNNs architectures used, implemented in Keras 2.10.0 (Chollet, 2015) over TensorFlow 2.10.0 (Abadi et al., 2016), were ResNet50 (He et al., 2016),

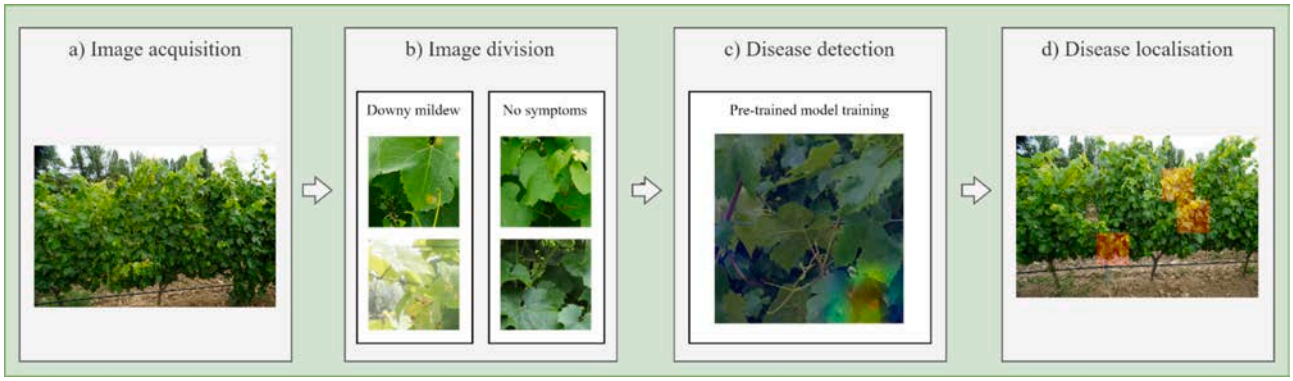

Fig. 1. Diagram of the methodology developed in this paper. Grapevine canopy images were acquired under field conditions (a). Images were divided in sub-images (b). Areas with downy mildew symptoms were detected training a pre-trained deep learning model and the model was verified using Grad-CAM (c). Areas with downy mildew symptoms were located in the full image using the trained model for object detection (d).

Table 1
Details of the plots at the time the images were taken and the number of images taken with each device.

Plot ID	Plot features		Image acquisition		Number of images captured			
	Grapevine variety	Downy mildew incidence	Time of day	Sky	On-the-go	Manually 1 (Canon)	Manually 2 (Sony)	Total
1	Hondarribi Zuri	Middle	Early afternoon	Cloudy	0	31	0	31
2	Tempranillo	Low	Early afternoon	Partly cloudy	0	0	8	8
3	Hondarribi Zuri	High	Early afternoon	Cloudy	0	12	10	22
4	Hondarribi Zuri Zerratia	High	Late afternoon	Cloudy	0	12	10	22
5	Hondarribi Zuri	High	Afternoon	Partly cloudy	0	11	10	21
6	Tempranillo	Very low	Early afternoon	Cloudy	0	0	3	3
7	Graciano	Low	Early afternoon	Cloudy	0	0	11	11
8	Tempranillo	Middle	Early evening	Clear	12	0	10	22
9	Hondarribi Zuri	High	Late morning	Cloudy	2	0	10	12
10	Hondarribi Zuri	High	Late morning	Cloudy	12	0	10	22
11	Hondarribi Zuri	Middle	Early afternoon	Cloudy	12	0	0	12
12	Hondarribi Zuri	Low	Afternoon	Partly cloudy	12	0	10	22
13	Hondarribi Zuri	Low	Late afternoon	Clear	2	0	5	7
14	Tempranillo	Low	Late afternoon	Clear	0	0	9	9
All					52	66	106	224

Fig. 2. Mobile platform used for taking the images on-the-go in the vineyard.

Fig. 3. Examples of images taken with different sky light conditions: clear (a), partly cloudy (b) and cloudy (c).

Xception (Chollet, 2017), EfficientNetV2S (Tan and Le, 2021) and RegNetX002 (Dollar et al., 2021). The vision transformer used, implemented in vit-keras library (Morales, 2021), was ViT-Base (Dosovitskiy et al., 2020) using patch sizes of 16 and 32 pixels. All models were trained using the weights of the pre-trained neural network: training only the classification layer, training only the normalization and classification layers (as in Frankle et al. (2020) for the ResNet architecture using ImageNet and CIFAR-10 datasets), and training all the layers. The neural networks minimized categorical cross entropy as the loss function using the Adam optimizer and a Cyclical Learning Rate (Smith, 2017)

Fig. 4. Examples of sub-images labelled as downy mildew (a) and non-downy mildew (b).

with an exponential function having a gamma value of 0.9, a step size of 1364 iterations (2*steps per epoch), a base learning rate of 0.001 and a maximum learning rate of 0.00001, that decrease the learning rate in each cycle. The batch size used to train the neural networks was 32. The training step stopped after 50 epochs without improving the validation accuracy and then, and the model achieving the best accuracy was selected.

All models were evaluated using a stratified hold-out validation, where the original image dataset, containing the high-resolution images, was separated into two distinct datasets: 80 % for training and validating the model, and 20 % for testing the model. The proportion of images from each plot was maintained in both datasets. Both the original images and the sub-images extracted from them were taken into consideration for the purposes of analysis. Subsequently, the sub-images from the dataset prepared for training and validating the model, were divided into two new datasets: one comprising 80 % of the sub-images, which were employed for model training, and the other comprising 20 %, which were utilised for model validation during the training process. The proportion of sub-images from each plot, the type of image acquisition (either on-the-go or manually using one of the two cameras), and the classification assigned by the expert were maintained. The total number of sub-images and images considered in each dataset is detailed in Table 2. The resulting labelling of the sub-images indicated that 43 % showed some downy mildew symptom and 57 % were areas without symptoms, maintaining a relatively balanced dataset. Online data augmentation was applied to the training dataset to improve the generalization ability during the training step. The sub-image transformations applied were the random combination of rotation between 0 and 360 degrees, horizontal and vertical flip, brightness change between 90 % and 110 % from its original brightness, zoom between the scales of 0.95 and 1.05, width and height shift between 0 and 5 % of its pixels and shearing between 0 and 5 degrees. In all cases, the points outside the boundaries of the input image were filled with the reflection of the image. The metrics used to evaluate the model classification were recall, precision, f1-score, accuracy, and confusion matrix. Their calculation formulas are as follows:

$$Recall = \frac{TP}{TP + FN} \tag{1}$$

Table 2
Number of sub-images (800x800 pixels) and images used in each dataset and separated by classes (downy mildew or no symptoms).

Dataset	Sub-images			Images
	Downy mildew	Non downy mildew	All	
Training	9584	12,258	21,842	179
Validation	2397	3065	5462	
Testing	2773	4047	6820	45
Total	14,754	19,370	34,124	224

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

The true positive values (TP) represented the number of sub-images correctly identified by the model as positive class, while the true negative values (TN) indicated the number of sub-images correctly classified as the negative class. The false positive values (FP) represented the number of negative instances incorrectly predicted as positive, and the false negative values (FN) referred to the number of positive instances incorrectly classified as negative. The recall, precision and f1-score metrics were calculated for each class, with the macro-average used for the final calculations.

Model prediction was visually analysed using two XAI methods, observing in both cases the pixels the neural network had prioritised to classify the images. The CNNs were analysed using the Gradient-weighted Class Activation Mapping (Grad-CAM, Selvaraju et al., 2020), using the gradient information from the last convolutional layer to assign importance values to the image pixels depending on a specific class. On the other side, the vision transformers were analysed using the output token of the network to see the attention associated with the input image, in the same way as Dosovitskiy et al. (2020).

3.4. Disease localisation

The sliding window method was used to analyse the complete canopy images and localise areas with disease symptoms. The trained neural network that achieved the highest f1-score was used to detect the windows where symptoms were present. The sliding window was 800x800 pixels and used a 400x400 pixel stride (50 % overlap), maintaining the same size used for the disease detection. The trained model classified the windows, identifying those that showed some of the symptoms. The windows classified as symptomatic of each image were then aggregated to enable a comprehensive evaluation of the entire image. This involved the segmentation of the image between the symptomatic areas and the areas where the disease was not detected. Furthermore, the same segmentation was conducted by reconstructing the images with the sub-images classified by the expert and aggregating those classified as symptomatic, thus enabling a comparison of the prediction with the expert's labels. The merged symptomatic areas were highlighted in red in the full image, thus facilitating the visualisation of the location of the areas containing disease symptoms. The localisation was evaluated with accuracy, f1-score, precision, recall and Intersection over Union (IoU). These metrics were evaluated by comparing the individual pixel values, taking into account the masks generated for each

image. Additionally, the severity of the disease on the plant was calculated as the percentage of the image that was identified as an area exhibiting symptoms of downy mildew. The root mean square error (RMSE) and coefficient of determination (R2) metrics were used to compare the severity obtained with the expert's labels with the predicted severity for each image. The metrics were calculated as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y}_i - y_i)^2}{n}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\widehat{y}_i - y_i)^2}{\sum_{i=1}^{n} (\overline{y} - y_i)^2} \quad (7)$$

Where y_i was the severity obtained with the expert assessment for the i-th full image, $\widehat{y}_i$ was the predicted severity for the i-th full image, $\overline{y}$ was the mean of severities of the expert and n the number of full images available in the analysed dataset.

The probability of a window being a diseased area was also evaluated by comparing the results of the location and severity prediction using threshold values between 0.1 and 0.9 on the probability. In this way, only windows whose probability of being diseased provided by the neural network was greater than the threshold were considered as diseased. The threshold that obtained a higher IoU for the images used in the training stage was taken into account to locate the symptoms of the disease.

4. Results and discussion

4.1. Downy mildew detection

Accurate detection of downy mildew or other plant diseases or pests is of great importance for effective agricultural management. The overall results of the classification process, as shown in Table 3, demonstrated the efficacy of the deep neural networks for identifying diseases such as downy mildew in grapevine. The analysis focused on testing different fine-tuning approaches in pre-trained neural network architectures to determine the optimal model. When the architecture is analysed independently of the fine-tuning, the vision transformer architectures, proposed by Dosovitskiy et al. (2020), produced the superior results. Under this condition, the features extracted by vision transformers using the ImageNet dataset seemed to be more general and useful than using CNNs, resulting in a significant increase in accuracy during the testing process, from 0.59 (using RegNet002) to 0.79 (using ViT-Base with a patch size of 16). In addition, fine-tuning seemed to improve the performance of all tested architectures, particularly in CNN architectures (Table 3). This reflects the utility of using fine-tuning with a pre-trained deep neural network using the ImageNet dataset, as it was demonstrated in the identification of rice (Chen et al., 2021) or apple (Li et al., 2021) leaf diseases. In line with Bevers et al., (2022), who conducted experiments testing various freezing approaches, the most favourable results were obtained retraining the neural networks from scratch. This was particularly evident in CNNs, where the EfficientNetV2S architecture, proposed by Chollet (2017), reached an accuracy of 0.91 and an F1-score of 0.92 in the testing phase. Similar results were obtained with the remaining CNNs, while ViT-Base indicated slighter lower performance. Vision transformer have demonstrated superior performance among state-of-the-art CNNs in other leaf diseases, like cassava's (Thai et al., 2021), but this is typically because of its large dataset and manual image acquisition. The identification of apple leaf diseases with complex backgrounds in the field was also attempted using these neural networks (Li et al., 2021), where the RegNet architecture was found to be more effective than ViT. This suggests that ViT could perform better considering a simple and extensive dataset. However, for identifying leaf diseases in complex scenarios, such as the one considered in this study that considers various daylight conditions and grapevine varieties, CNNs may still be a more suitable solution. It is noteworthy that retraining only the normalization layers appeared to be a more robust approach than retraining the models from scratch, achieving similar results in the validation and testing phases and requiring fewer training layers. This may be attributed to the limited amount of data available, which could result in overfitting of the complex models. This suggest that the EfficientNetV2S architecture could be adapted effectively to a small dataset, although for other models retraining only part of the layers could help to reduce overfitting.

The application of XAI through the Grad-CAM method facilitated the interpretation of the results produced by all CNN architectures. The Grad-CAM method enabled the visualisation of the specific areas of the images that the neural networks focused on for classification. In the examples presented in Fig. 5, it can be seen that all CNNs had focused on the areas of leaves with yellow spots (oil spots characteristic of leaves infected by *Plasmopara viticola*), to detect leaves with downy mildew symptoms. On the other hand, in the detection of the non-infected areas, each network seems to concentrate on areas of the plant where parts of the leaves were present, yet no such oil spots were discernible. This visualization corroborated the correct functioning of CNNs highlighting image regions that distinguish labels, filling the gap of interpretation in deep neural networks for leaf disease detection, as it is commented in Liu and Wang (2021).

In the case of vision transformer, the results of the neural networks

Table 3
Results of the classification of the sub-images on the validation and test set with all the models. The best model is highlighted in bold.

Fine Tuning	Trained layers	Architecture	Validation				Testing			
			Accuracy	F1-score	Precision	Recall	Accuracy	F1-score	Precision	Recall
None	1	EfficientNetV2S	0.63	0.64	0.70	0.60	0.61	0.63	0.73	0.56
	1	RegNetX002	0.56	0.72	0.56	1.00	0.59	0.74	0.59	1.00
	1	ResNet50	0.64	0.68	0.68	0.68	0.62	0.68	0.69	0.67
	1	Vit-Base16	0.80	0.83	0.80	0.86	0.79	0.82	0.82	0.83
	1	Vit-Base32	0.78	0.82	0.78	0.85	0.78	0.81	0.81	0.82
	1	Xception	0.75	0.79	0.76	0.81	0.73	0.78	0.77	0.79
Normalization layers	111	EfficientNetV2S	0.89	0.90	0.92	0.89	0.89	0.91	0.93	0.89
	45	RegNetX002	0.87	0.89	0.88	0.89	0.86	0.88	0.89	0.87
	54	ResNet50	0.87	0.89	0.89	0.89	0.86	0.88	0.90	0.87
	2	Vit-Base16	0.79	0.82	0.80	0.85	0.79	0.82	0.82	0.82
	2	Vit-Base32	0.78	0.81	0.79	0.84	0.77	0.81	0.81	0.81
	41	Xception	0.89	0.90	0.90	0.90	0.88	0.89	0.91	0.88
All layers	**515**	**EfficientNetV2S**	**0.92**	**0.93**	**0.94**	**0.92**	**0.91**	**0.92**	**0.94**	**0.90**
	145	RegNetX002	0.90	0.91	0.92	0.90	0.87	0.89	0.92	0.86
	177	ResNet50	0.92	0.93	0.94	0.92	0.90	0.91	0.94	0.89
	20	Vit-Base16	0.88	0.89	0.90	0.89	0.85	0.87	0.89	0.85
	20	Vit-Base32	0.86	0.87	0.88	0.87	0.77	0.80	0.83	0.78
	134	Xception	0.92	0.93	0.93	0.92	0.90	0.91	0.94	0.89

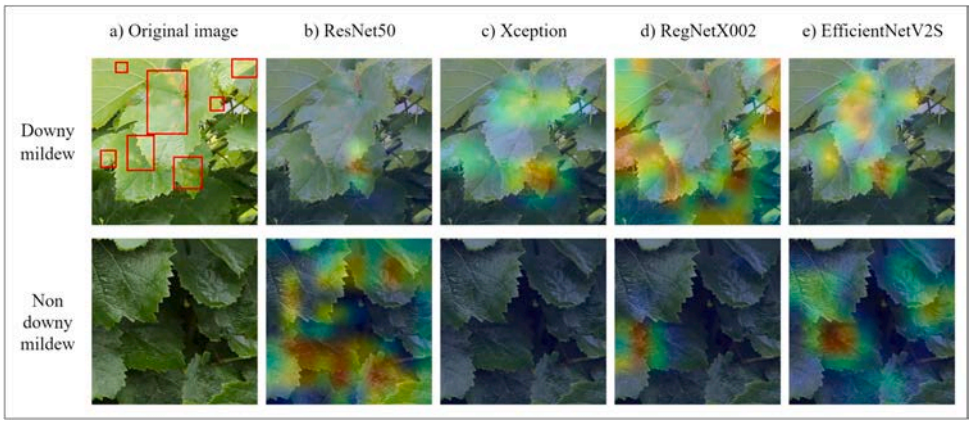

Fig. 5. Examples of Grad-CAM results using the best models of each CNN architecture in windows with and without downy mildew symptoms. Regions with downy mildew symptoms are marked in the original image by red rectangles.

were analysed using attention maps, as can be seen in the Fig. 6, in a similar way as Chen et al. (2021) did to analyse the rice disease classification. The attention maps revealed that the trained models were focusing on the oil spots in the diseased leaves and on other regions in the non-diseased leaves. In addition, the main difference between the two vision transformers was the certainty in predicting areas with downy mildew symptoms. The Vit-Base16 paid more attention to the oil spots than the rest of the image. This discrepancy may be attributed to the utilisation of an architecture with a smaller path size for the extraction of important features from the images. The employment of a path size of 16 pixels facilitated a more detailed and fine-grained focus, whereas a path of 32 pixels offered a more generalised view. In this context, the Vit-Base16 architecture yielded more precise outcomes, enabling the focus of the detection on the oil spots.

Considering all the maps obtained with the XAI methods, it could be considered that all the architectures were focusing on the parts of the image containing disease symptoms to classify them. The ability of these methods to interpret the prediction using deep learning was also demonstrated, avoiding the need to extract features by hand (Araujo and Peixoto, 2019; Wu et al., 2020) or to analyse feature maps of convolutional layers individually (Arumuga Arun and Umamaheswari, 2023). Moreover, the fully retrained EfficientNetV2S architecture reflected its good performance with the Grad-CAM heatmap, localising most of the oil spots present in the diseased image.

The results of the top-performing model, retraining all the layers of the EfficientNetV2S architecture, were analysed using a confusion matrix (Table 4). The study indicates that a significant proportion of true positives (TP) and true negatives (TN), representing 86 % and 94 % of the positive and negative samples, respectively, was detected. These results showed a favourable balance between the two classes, resulting in high precision for both cases. Another notable value in the confusion matrix was the number of false positives (FP), indicating the amount of non-symptomatic areas misclassified as symptomatic. Only 6 % of the negative samples were wrongly classified in this category. Some examples of such false positives using the Grad-CAM method are shown in Fig. 7. These examples seemed to reflect that false positives were mainly due to leaf damages with similar features to downy mildew symptoms, so their detection could be positive in order to analyse these damages in depth and avoid ignoring real symptoms. The utilisation of images captured in complex field conditions, where there may be vegetation on the ground, plants in the background, or diverse lightning conditions, may potentially impact the precision of the model. Nevertheless, the developed methodology enables the examination of these potential shortcomings and could be enhanced with new sub-images containing these conflicting attributes through straightforward labelling.

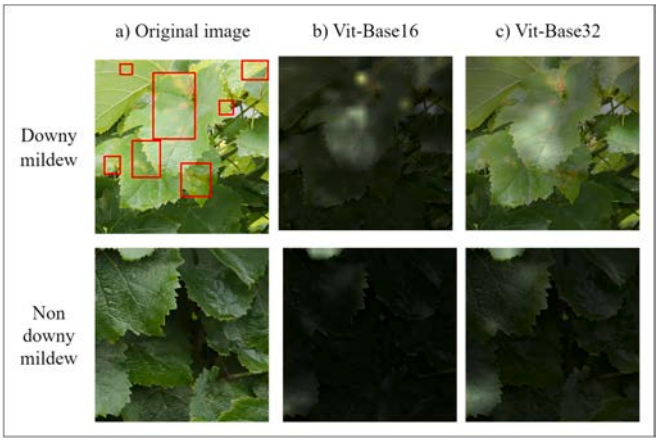

Fig. 6. Examples of attention map results using the best models of each Vit architecture in windows with and without downy mildew symptoms. Regions with downy mildew symptoms are marked in the original image by red rectangles.

4.2. Downy mildew localisation

Once the best model was trained, the localisation of downy mildew symptoms was completed using a sliding window. The thresholds tested to indicate whether a downy mildew detection was considered based on the probability provided by the neural network are summarized in Table 5 and Fig. 8. It can be seen that the localisation results were better using a threshold of 0.6, obtaining an IoU of 0.83 and a F1-score of 0.89. In addition, the results comparing the percentage of the image that contain downy mildew symptoms detected by an expert and automatically were similar for all the thresholds, obtaining the lowest RMSE and the highest R^2 for the thresholds between 0.4 and 0.7. This indicates that the medium thresholds are more accurate in locating downy mildew symptoms. For this reason and considering that localisation was performed with a 50 % overlap, which helped to reinforce symptom detection, a 0.6 threshold was used to detect the areas of the grapevine canopy with downy mildew symptoms. This method allowed supervised

Table 4
Confusion matrix of the classification of the windows from the test set using the EfficientNetV2S architecture retraining all its layers.

		Actual class	
		Downy mildew	Non downy mildew
Predicted class	Downy mildew	2542 (86 %)	231 (6 %)
	Non downy mildew	414 (14 %)	3633 (94 %)

Fig. 7. Grad-CAM examples for False positive values using the EfficientNetV2S architecture training all its layers.

Table 5
Results of the localisation of the areas with downy mildew symptoms for the full images on the test set with the best model. The location of the areas was compared using the IoU, accuracy, F1-score, precision, and recall metrics. The best threshold is highlighted in bold.

Threshold	Training and validation					Testing				
	IoU	Accuracy	F1	Precision	Recall	IoU	Accuracy	F1	Precision	Recall
0.1	0.83	0.90	0.89	0.83	0.99	0.74	0.82	0.83	0.74	0.98
0.15	0.86	0.93	0.91	0.87	0.98	0.76	0.85	0.85	0.78	0.97
0.2	0.88	0.94	0.93	0.89	0.98	0.78	0.87	0.86	0.80	0.96
0.25	0.89	0.95	0.93	0.90	0.98	0.79	0.88	0.87	0.82	0.96
0.3	0.91	0.95	0.94	0.92	0.97	0.80	0.88	0.87	0.83	0.95
0.35	0.91	0.96	0.95	0.93	0.97	0.81	0.89	0.88	0.85	0.94
0.4	0.92	0.96	0.95	0.94	0.96	0.82	0.90	0.89	0.87	0.93
0.45	0.92	0.96	0.95	0.94	0.96	0.82	0.90	0.88	0.87	0.92
0.5	0.92	0.96	0.95	0.95	0.95	0.82	0.90	0.89	0.88	0.92
0.55	0.92	0.96	0.95	0.96	0.95	0.83	0.91	0.89	0.89	0.91
0.6	**0.92**	**0.96**	**0.95**	**0.96**	**0.94**	**0.83**	**0.91**	**0.89**	**0.90**	**0.90**
0.65	0.92	0.96	0.95	0.97	0.94	0.83	0.91	0.89	0.91	0.90
0.7	0.91	0.96	0.95	0.97	0.93	0.83	0.91	0.89	0.92	0.89
0.75	0.91	0.96	0.94	0.97	0.92	0.82	0.91	0.89	0.92	0.88
0.8	0.90	0.95	0.94	0.98	0.91	0.82	0.91	0.89	0.93	0.87
0.85	0.89	0.95	0.93	0.98	0.89	0.81	0.91	0.88	0.94	0.85
0.9	0.87	0.94	0.92	0.98	0.87	0.79	0.90	0.87	0.95	0.83

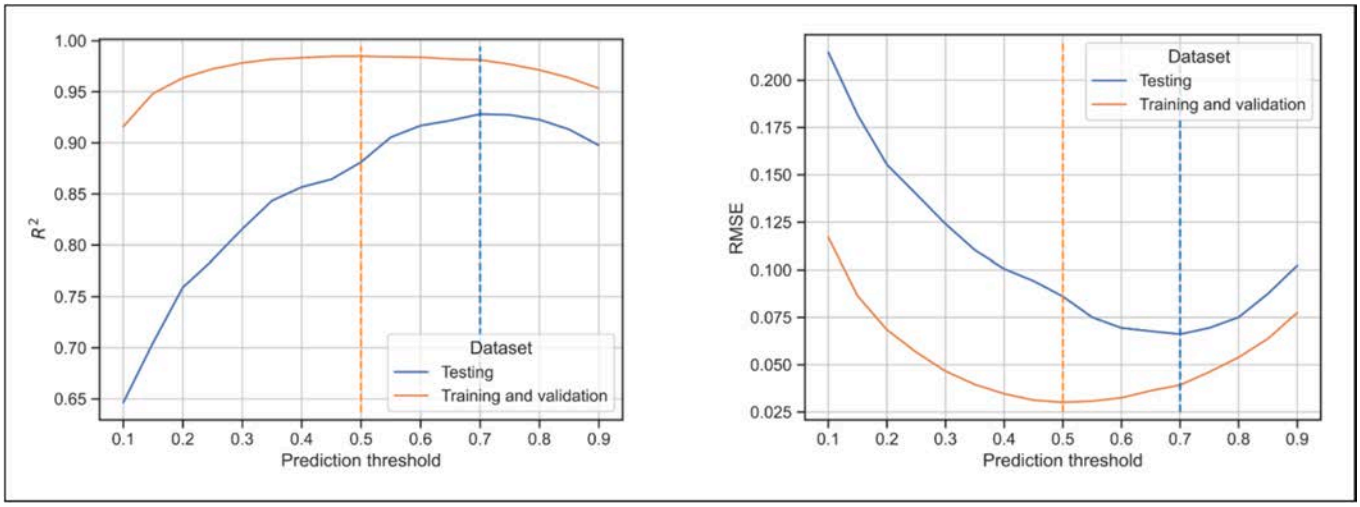

Fig. 8. Results of the localisation of the areas with downy mildew symptoms for the full images on the test set with the best model. The percentage of the area in the image with downy mildew symptoms was compared with the expert detection using the R2 and RMSE metrics. The dashed vertical lines indicate the best threshold for each dataset.

symptom localisation throughout the entire plant, even in the cases where spots are hardly visible. It accomplished this by avoiding the need for time-consuming labelling of disease lesion (Gao et al., 2021; Zhang et al., 2022), and instead taking into account the context of the infected area. Additionally, it reduced the false positives due to small plant damage areas, dark leaves or soil patches.

An example of the localisation of downy mildew symptoms in a full image can be seen in the Fig. 9. This type of image enables consideration of the location of the downy mildew infection on the leaf surface, in addition to its detection. In this example, all symptoms of downy mildew were successfully identified, although some areas were discarded, as they were covered by other overlapping windows. The complexity of the high-quality images used is also remarkable, reflecting the whole plant and some elements that could be conflicting, such as the ground, sky or rows of the same crop in the background. Such images also indicate the potential for utilising ground vehicles, such as the platform used in this work, tractors or other crop-specific vehicles, for disease assessment purposes. The vehicles can be equipped with imaging systems to capture comprehensive views of the plants, thereby facilitating precise disease detection across extensive crop fields. This reflects the advantage of this method to detect the disease without having to focus the analysis on parts of the plant susceptible to symptoms (Chen et al., 2021; Kumar Sahu and Pandey, 2023; Wu et al., 2020), and in real field conditions. In addition, these results could indicate the possibility of using this method to detect low-infected areas of the crop, allowing a more accurate treatment of the area.

4.3. Method generalization

The RGB images were taken under different conditions, using different cameras, on-the-go or manually, under different lighting conditions (mainly based on agrometeorological conditions) and on different grape varieties. This diversity contributed to the complexity of the dataset, leading to the development of a robust model adaptable to real-world scenarios. As indicated in Figure 10, the method adjusted accordingly to variations in acquiring the images and lighting conditions. The results presented F1-score values ranging between 0.85 and 0.93 in disease detection and IoU values between 78 and 91 % in disease localisation. The findings suggest that the method could adapt to a mobile platform comparably to manual image acquisition. Therefore, it may be feasible to implement it in the vehicles used in cultivation, as in Abdelghafour et al. (2020) work. Similarly, obtained results from images captured under various sky conditions indicated good performance of technology in natural conditions. The need to take the images at night or under cloudy skies (Tardif et al., 2022) to avoid noise caused by heterogeneous day-light or the influence of plants in the background is not necessary. The grapevine varieties showed similar results, obtaining the best results for the Hondarribi Zuri variety, most likely due to the availability of a greater number of images. On the other hand, the Graciano variety obtained lower F1-score and IoU values, as a result of considering fewer images. Thus, for effective training of the neural network, the method developed seemed to work better when more than 9 full images per condition are included, as in the case of grapevine variety. While adaptable, the performance of the method is tied to the conditions under which it was trained, meaning it might need retraining or fine-tuning when applied to significantly different conditions or crop types. Overall, this method demonstrated significant potential for adaptation to other grape varieties or conditions, requiring minimal data to achieve satisfactory results, and could be implemented in large crops with similar conditions.

Additionally, the figure illustrates the discrepancy between the severity estimated by the expert and that predicted by the method, calculated from the disease localisation. At the highest severities an almost negligible error was made, while at the lowest severities the error

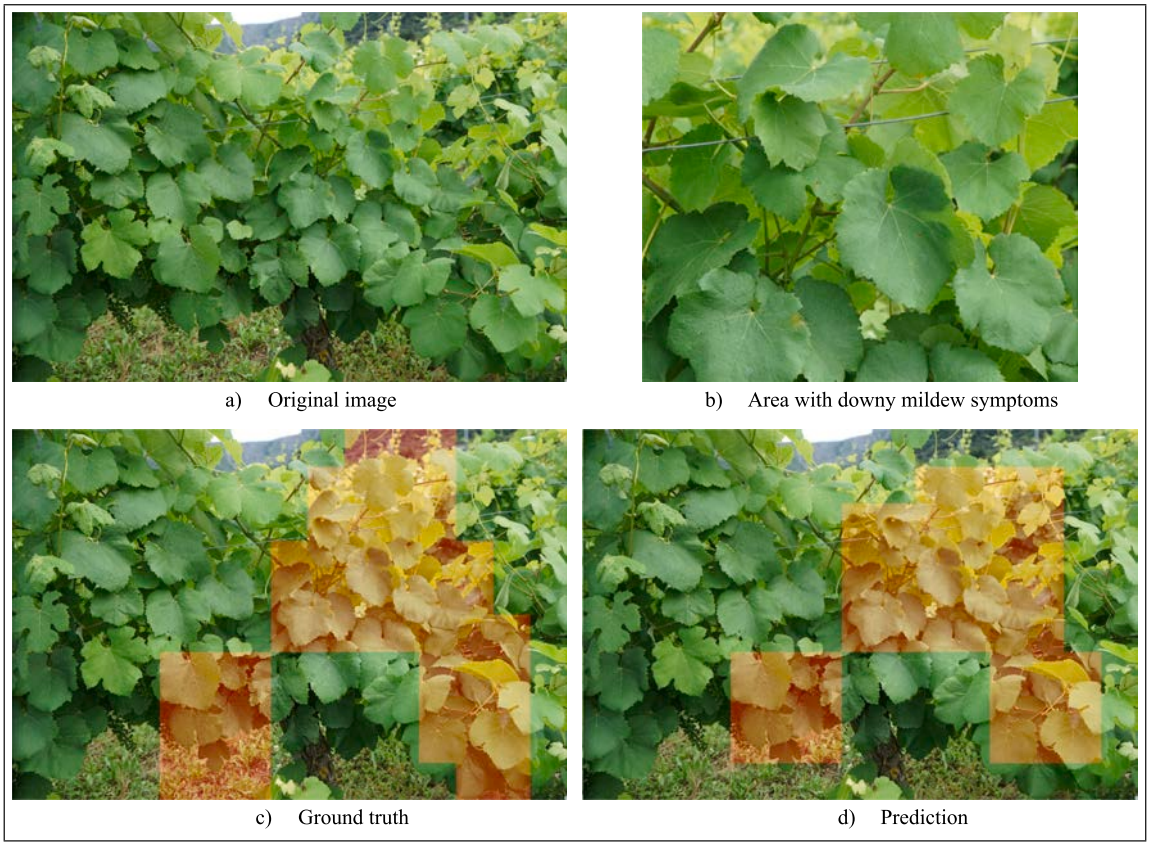

Fig. 9. Example of image taken in the field (a), one of the areas with downy mildew symptoms (b), downy mildew localisation by an expert (a) and using the EfficientNet architecture retraining all the layers and applying a detection threshold of 0.6 (b).

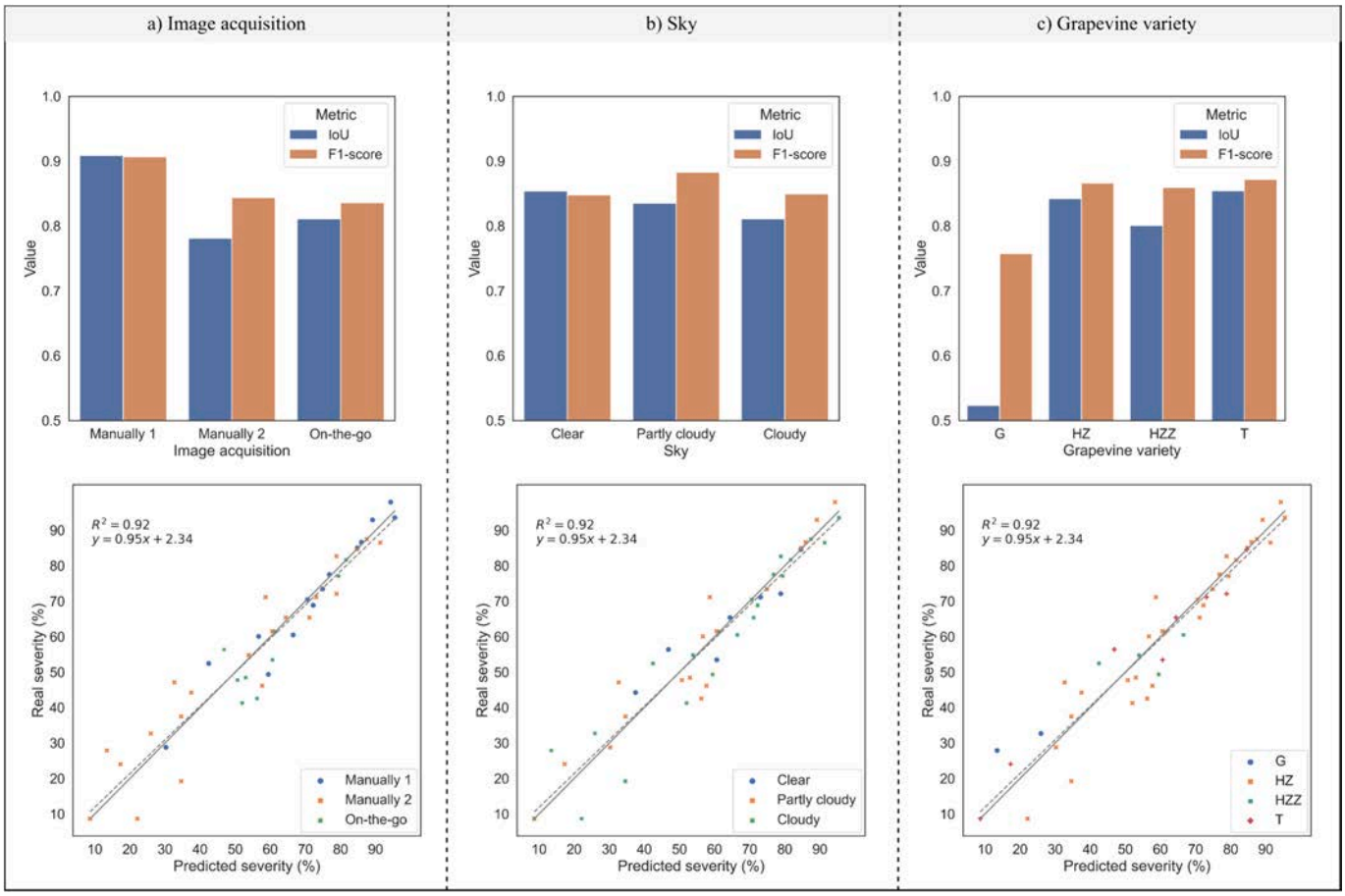

Fig. 10. Results of downy mildew detection (evaluating the F1 score metric) and localisation (evaluating the Intersection over Union metric) separated by conditions (image acquisition, sky or grapevine variety) using the test dataset. The images of each condition are separated by the colour and shape associated to its condition in the scatterplot showing the relationship between real and predicted severity of each image. The solid lines represent the perfect concordance line. The dashed lines indicate the actual tendency of the relationship. The manual image acquisition was divided into two categories: manually 1, using a Canon camera, and manually 2, using a Sony camera. The grapevine varieties analysed were Graciano (G), Tempranillo (T), Hondarribi Zuri (HZ) and Hondarribi Zuri Zerratia (HZZ).

increased. In the latter case, the method seemed to tend to predict more severity than the existing one, although without departing too far from the general trend. This characteristic may be particularly advantageous for the early detection of the disease, allowing for timely adjustments to treatments in less infected areas and thereby reducing critical damage to the crop. This quantification capability might enhance crop management by enabling more precise and effective disease intervention strategies.

5. Conclusions

This study focused on the identification and localisation of diseases in real field conditions using an interpretable deep learning approach that is readily adaptable to new conditions through the use of transfer learning and straightforward data labelling. The case study employed was that the detection of grapevine downy mildew in entire plants. The neural networks were retrained with several fine-tuning approaches, resulting in an optimal balance between model robustness and accuracy by retraining only the normalisation layers of the neural network. Furthermore, the EfficientNetV2S architecture achieved an accuracy of 91 % in the test dataset upon retraining all of its layers, thereby demonstrating its capability to adapt efficiently to small complex datasets. In general, CNNs proved to be superior to ViT in adapting to complex and small datasets. Moreover, the developed method facilitated the examination of neural network predictions using XAI techniques, revealing that the models concentrated on small oil spots for disease detection, particularly the optimal model, which automatically identified the majority of the visible symptoms. The techniques demonstrated high performance on a modest dataset of 179 high-quality images, indicating the potential applicability to other grapevine varieties, pests/diseases or crops. The method allowed rapid and precise localisation of downy mildew in images of the grapevine canopy, even when symptoms were minimal, reducing the need for arduous image annotation. Additionally, the method demonstrated adaptability to different grapevine varieties, automatically identifying similarities among their downy mildew symptoms. Moreover, the results were consistent for images captured under various daylight conditions, whether using a mobile platform or a static setup. This indicates the possibility of integrating this algorithm into crop treatment machines, such as tractors. Such integration could reduce environmental impact by allowing variable-rate applications of pesticides. This algorithm opens the possibility for accurate real-time crop monitoring and the prevention of major disease damage, marking significant advancements in agricultural disease management.

CRediT authorship contribution statement

Inés Hernández: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Salvador Gutiérrez:** Methodology, Investigation, Formal analysis. **Ignacio Barrio:** Validation, Methodology, Formal analysis, Data curation. **Rubén Íñiguez:** Validation, Methodology, Formal analysis, Data curation. **Javier**

Tardaguila: Writing – original draft, Validation, Supervision, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Data availability

The data that has been used is confidential.

Acknowledgements

This work has been developed as part of the project NoPest (Novel Pesticides for a Sustainable Agriculture), which received funding from the European Union Horizon 2020 FET Open program under Grant agreement ID 828940. Inés Hernández and Rubén Iñíguez would like to acknowledge the research funding FPI PhD grants 1150/2020 and 591/2021 by Universidad de La Rioja and Gobierno de La Rioja.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016 . TensorFlow: Large-scale machine learning on heterogeneous distributed systems. ArXiv: 10.48550/arXiv.1603.04467.

Abdelghafour, F., Keresztes, B., Germain, C., Da Costa, J.-P., 2020. In field detection of downy mildew symptoms with proximal colour imaging. Sensors 20, 4380. https://doi.org/10.3390/s20164380.

Agarwal, M., Gupta, S.K., Biswas, K.K., 2020. Development of Efficient CNN model for Tomato crop disease identification. Sustain. Comput. Informatics Syst. 28, 100407. https://doi.org/10.1016/j.suscom.2020.100407.

Ahmad, A., Saraswat, D., El Gamal, A., 2023. A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools. Smart Agric. Technol. 3, 100083. https://doi.org/10.1016/j.atech.2022.100083.

Al-Saddik, H., Laybros, A., Billiot, B., Cointault, F., 2018. Using image texture and spectral reflectance analysis to detect Yellowness and Esca in grapevines at leaf-level. Remote Sens. 10, 618. https://doi.org/10.3390/rs10040618.

Araujo, J.M.M., Peixoto, Z.M.A., 2019. A new proposal for automatic identification of multiple soybean diseases. Comput. Electron. Agric. 167, 105060. https://doi.org/10.1016/j.compag.2019.105060.

Arumuga Arun, R., Umamaheswari, S., 2023. Effective multi-crop disease detection using pruned complete concatenated deep learning model. Expert Syst. Appl. 213, 118905. https://doi.org/10.1016/j.eswa.2022.118905.

Barbedo, J.G.A., 2014. An automatic method to detect and measure leaf disease symptoms using digital image processing. Plant Dis. 98, 1709–1716. https://doi.org/10.1094/PDIS-03-14-0290-RE.

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Bevers, N., Sikora, E.J., Hardy, N.B., 2022. Soybean disease identification using original field images and transfer learning with convolutional neural networks. Comput. Electron. Agric. 203, 107449. https://doi.org/10.1016/j.compag.2022.107449.

Cai, C., Wang, Q., Cai, W., Yang, Y., Hu, Y., Li, L., Wang, Y., Zhou, G., 2023. Identification of grape leaf diseases based on VN-BWT and Siamese DWOAM-DRNet. Eng. Appl. Artif. Intell. 123, 106341. https://doi.org/10.1016/j.engappai.2023.106341.

Chen, Y., Wu, Q., 2023. Grape leaf disease identification with sparse data via generative adversarial networks and convolutional neural networks. Precis. Agric. 24, 235–253. https://doi.org/10.1007/s11119-022-09941-z.

Chen, J., Zhang, D., Zeb, A., Nanehkaran, Y.A., 2021. Identification of rice plant diseases using lightweight attention networks. Expert Syst. Appl. 169, 114514. https://doi.org/10.1016/j.eswa.2020.114514.

Chollet, F., 2015. Keras.

Chollet, F., 2017. Xception: deep learning with depthwise separable convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

Dey, B., Masum Ul Haque, M., Khatun, R., Ahmed, R., 2022. Comparative performance of four CNN-based deep learning variants in detecting Hispa pest, two fungal diseases, and NPK deficiency symptoms of rice (Oryza sativa). Comput. Electron. Agric. 202, 107340. doi: 10.1016/j.compag.2022.107340.

Dollar, P., Singh, M., Girshick, R., 2021. Fast and accurate model scaling, in: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 924–932. doi: 10.1109/CVPR46437.2021.00098.

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. ArXiv. https://doi.org/10.48550/arXiv.2010.11929.

Frankle, J., Schwab, D.J., Morcos, A.S., 2020. Training BatchNorm and only BatchNorm: On the expressive power of random features in CNNs. ArXiv: 10.48550/arXiv.2003.00152.

Gao, J., Westergaard, J.C., Sundmark, E.H.R., Bagge, M., Liljeroth, E., Alexandersson, E., 2021. Automatic late blight lesion recognition and severity quantification based on field imagery of diverse potato genotypes by deep learning. Knowledge-Based Syst. 214, 106723. https://doi.org/10.1016/j.knosys.2020.106723.

Gessler, C., Pertot, I., Perazzolli, M., 2011. Plasmopara viticola: A review of knowledge on downy mildew of grapevine and effective disease management. Phytopathol. Mediterr. 50, 3–44. https://doi.org/10.14601/Phytopathol_Mediterr-9360.

Gutierrez, S., Hernandez, I., Ceballos, S., Barrio, I., Diez-Navajas, A.M., Tardaguila, J., 2021. Deep learning for the differentiation of downy mildew and spider mite in grapevine under field conditions. Comput. Electron. Agric. 182, 105991. https://doi.org/10.1016/j.compag.2021.105991.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 770–778. doi: 10.1109/CVPR.2016.90.

Hernandez, I., Gutierrez, S., Ceballos, S., Iñiguez, R., Barrio, I., Tardaguila, J., 2021. Artificial intelligence and novel sensing technologies for assessing downy mildew in grapevine. Horticulturae 7, 103. https://doi.org/10.3390/horticulturae7050103.

Hernandez, I., Gutierrez, S., Ceballos, S., Palacios, F., Toffolatti, S.L., Maddalena, G., Diago, M.P., Tardaguila, J., 2022. Assessment of downy mildew in grapevine using computer vision and fuzzy logic. Development and validation of a new method. OENO One 56, 41–53. https://doi.org/10.20870/oeno-one.2022.56.3.5359.

Kakogeorgiou, I., Karantzalos, K., 2021. Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. Int. J. Appl. Earth Obs. Geoinf. 103, 102520. https://doi.org/10.1016/j.jag.2021.102520.

Kumar Sahu, S., Pandey, M., 2023. An optimal hybrid multiclass SVM for plant leaf disease detection using spatial Fuzzy C-Means model. Expert Syst. Appl. 214, 118989. https://doi.org/10.1016/j.eswa.2022.118989.

Kusrini, K., Suputa, S., Setyanto, A., Agastya, I.M.A., Priantoro, H., Chandramouli, K., Izquierdo, E., 2020. Data augmentation for automated pest classification in mango farms. Comput. Electron. Agric. 179, 105842. https://doi.org/10.1016/j.compag.2020.105842.

Lee, W.S., Tardaguila, J., 2023. Pest and disease management, in: Advanced Automation for Tree Fruit Orchards and Vineyards. Springer, Cham, pp. 93–118. doi: 10.1007/978-3-031-26941-7_5.

Li, L., Zhang, S., Wang, B., 2021. Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. Sensors 22, 173. https://doi.org/10.3390/s22010173.

Liu, E., Gold, K.M., Combs, D., Cadle-Davidson, L., Jiang, Y., 2022. Deep semantic segmentation for the quantification of grape foliar diseases in the vineyard. Front. Plant Sci. 13, 3342. https://doi.org/10.3389/fpls.2022.978761.

Liu, J., Wang, X., 2021. Plant diseases and pests detection based on deep learning: a review. Plant Methods 17, 22. https://doi.org/10.1186/s13007-021-00722-9.

Mastrodimos, N., Lentzou, D., Templalexis, C., Tsitsigiannis, D.I., Xanthopoulos, G., 2019. Development of thermography methodology for early diagnosis of fungal infection in table grapes: The case of Aspergillus carbonarius. Comput. Electron. Agric. 165, 104972. https://doi.org/10.1016/j.compag.2019.104972.

Morales, F., 2021. vit-keras.

Nguyen, C., Sagan, V., Maimaitiyiming, M., Maimaitijiang, M., Bhadra, S., Kwasniewski, M.T., 2021. Early detection of plant viral disease using hyperspectral imaging and deep learning. Sensors 21, 742. https://doi.org/10.3390/s21030742.

Paulus, I., De Busscher, R., Schrevens, E., 1997. Use of image analysis to investigate human quality classification of apples. J. Agric. Eng. Res. 68, 341–353. https://doi.org/10.1006/jaer.1997.0210.

Pintelas, E., Liaskos, M., Livieris, I.E., Kotsiantis, S., Pintelas, P., 2021. A novel explainable image classification framework: case study on skin cancer and plant disease prediction. Neural Comput. Appl. 33, 15171–15189. https://doi.org/10.1007/S00521-021-06141-0.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. Int. J. Comput. vis. 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2020. Grad-CAM: Visual explanations from deep networks via gradient-based localization. Int. J. Comput. vis. 128, 336–359. https://doi.org/10.1007/s11263-019-01228-7.

Smith, L.N., 2017. Cyclical learning rates for training neural networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, pp. 464–472. doi: 10.1109/WACV.2017.58.

Tan, M., Le, Q.V., 2021. EfficientNetV2: Smaller models and faster training. ArXiv. https://doi.org/10.48550/arXiv.2104.00298.

Tardif, M., Amri, A., Keresztes, B., Deshayes, A., Martin, D., Greven, M., Da Costa, J.P., 2022. Two-stage automatic diagnosis of Flavescence Dorée based on proximal

imaging and artificial intelligence: a multi-year and multi-variety experimental study. Oeno One 56, 371–384. https://doi.org/10.20870/oeno-one.2022.56.3.5460.
Thai, H.-T., Tran-Van, N.-Y., Le, K.-H., 2021. Artificial cognition for early leaf disease detection using vision transformers, in: 2021 International Conference on Advanced Technologies for Communications (ATC). IEEE, pp. 33–38. doi: 10.1109/ATC52653.2021.9598303.
Thakur, P.S., Khanna, P., Sheorey, T., Ojha, A., 2022. Trends in vision-based machine learning techniques for plant disease identification: A systematic review. Expert Syst. Appl. 208, 118117. https://doi.org/10.1016/j.eswa.2022.118117.
Wilcox, W.F., Gubler, W.D., Uyemoto, J.K., 2015. Compendium of grape diseases, disorders, and pests, second edition. The American Phytopathological Society. doi: 10.1094/9780890544815.
Wu, A., Zhu, J., He, Y., 2020. Computer vision method applied for detecting diseases in grape leaf system. Studies in Computational Intelligence. 367–376. https://doi.org/10.1007/978-3-030-04946-1_36.
Xia, X., Chai, X., Zhang, N., Sun, T., 2021. Visual classification of apple bud-types via attention-guided data enrichment network. Comput. Electron. Agric. 191, 106504. https://doi.org/10.1016/j.compag.2021.106504.
Zhang, Z., Qiao, Y., Guo, Y., He, D., 2022. Deep learning based automatic grape downy mildew detection. Front. Plant Sci. 13, 1–12. https://doi.org/10.3389/fpls.2022.872107.
Zhou, C., Zhong, Y., Zhou, S., Song, J., Xiang, W., 2023. Rice leaf disease identification by residual-distilled transformer. Eng. Appl. Artif. Intell. 121, 106020. https://doi.org/10.1016/j.engappai.2023.106020.