



Fusion architectures for soft rot detection in melon plants using hyperspectral and multicolor fluorescence imaging

Salvador Gutiérrez ^{a,*}, Matilde Barón ^b, María Trinidad Moreno-Martín ^b,
Mónica Pineda ^{b,*}

^a Department of Computer Science and Artificial Intelligence (DECSAI), Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada (UGR), Granada, 18014, Spain

^b Departamento de Estrés, Desarrollo y Señalización en Plantas, Estación Experimental del Zaidín (EEZ-CSIC), Granada, 18008, Spain

ARTICLE INFO

Keywords:

B. cinerea
Deep learning
Sensor fusion
Hyperspectral imaging
Multicolor fluorescence
Early disease detection
Precision agriculture

ABSTRACT

The wide range of non-destructive sensors currently available in agriculture encourages the development of fusion models for better monitoring tools than individual devices. In this work, two deep-learning fusion architectures that combine hyperspectral and multicolor fluorescence imaging were designed and tested for the early detection of soft rot caused by *Botrytis cinerea* in melon (*Cucumis melo*) leaves, and compared them with four single-sensor architectures. Melon leaf sampling yielded 739 samples (431 infected, 308 healthy). Each sample combined an hyperspectral imaging spectrum (400–1000 nm, 300 bands) with four fluorescence images taken at 440, 520, 680, and 740 nm. Sampling was performed 1 and 2 days post-inoculation (dpi), when no visual symptoms were present, to train deep-learning models discriminating between infected and healthy plants. The six architectures were evaluated by training a large number of models, resulting from a combination of 48 hyperparameter fits and repeated 5-fold cross-validation, for a total of 4320 models. Statistical analyses showed that one fusion architecture significantly outperformed all others, achieving the highest mean cross-validated accuracy (0.8068), with individual runs reaching a maximum accuracy of 0.8919 for the significantly better-performing fusion architecture. These results confirm that the combination of reflectance spectra with four channels of fluorescence improves early soft rot detection in melon compared with single-sensor solutions.

1. Introduction

Today, there are many sensing devices available, with solutions for both small- and large-scale customers. This number is constantly increasing as technology develops and costs decrease. In agriculture, sensing technologies are assuming a progressively crucial function throughout all operational phases [1–3], with reported applications for crop phenotyping [4], growth and development control [5] or leaf and fruit monitoring [6–8]. Moreover, automated detection of plant diseases can be addressed by different types of sensors [9,10]. Thus, non-invasive monitoring of plant diseases can prevent their spread and further damage to crops, increasing yields and reducing the use of pesticides and their environmental impact [11].

In light of their remarkable capacity for discerning the physiological alterations occurring in infected plants even in the absence of symptoms, two sensors have recently been recognized as pivotal instruments for disease detection and plant phenotyping: multicolor fluorescence (MCF) and hyperspectral reflectance devices [12–16]. Thus, MCF imag-

ing (MCFI) has been successfully applied to detect several plant diseases in bean [17], melon [18], zucchini [19] and oilseed rape plants [20]; whereas spectral reflectance has been successful in grapevine [21], sugar beet [22], wheat [23], tomato [24], broccoli [7] and avocado [25]. On one hand, MCF devices capture the light emitted by plants at different wavelengths when illuminated with ultraviolet (UV) light. Fluorescence in the blue (F440) and green (F520) regions of the spectrum is emitted by phenolic compounds related to stress defence mechanisms; while red (F680) and far-red (F740) fluorescence is emitted by chlorophyll *a*. On the other hand, hyperspectral cameras capture high-resolution data, as two spatial dimensions form one image; each pixel is composed of a full spectrum of some wavelength range of interest.

As sensing technologies become more integrated into agricultural practices, the focus is shifting towards the effective application of the diverse data they produce. While individual sensors provide valuable insights, combining data from multiple sources could enable more accurate and reliable plant disease detection [26]. This opens the door to advanced data integration methods that can leverage the strengths of

* Corresponding authors.

E-mail addresses: salvaguti@decsai.ugr.es (S. Gutiérrez), matilde.baron@eez.csic.es (M. Barón), trini.moreno@eez.csic.es (M.T. Moreno-Martín), monica.pineda@eez.csic.es (M. Pineda).

<https://doi.org/10.1016/j.atech.2026.101892>

Received 20 October 2025; Received in revised form 17 January 2026; Accepted 15 February 2026

Available online 18 February 2026

2772-3755/© 2026 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

different sensors to enhance the performance of the entire system. In this context, the question arises as to whether multi-sensor data fusion provides better performance than single-sensor approaches. Therefore, there is an opportunity to explore the potential of deep learning to develop fusion architectures that comply with disease detection and perform substantially better than models trained from each sensing device. Recent work in agricultural imaging has shown that lightweight fusion architectures can achieve high performance in crop disease detection and yield prediction tasks [27,28], highlighting the importance of balancing model complexity with generalization performance in agricultural imaging applications. Nevertheless, the application of these methods must be optimized to find the most effective models possible with the shortest computing times, which implies the advantage of greater algorithms efficiency with lower energy consumption [29]. Such is the case with convolutional neural networks (CNNs) and multilayer perceptrons (MLPs) [30].

Data fusion has been applied to agriculture to some extent, most extensively in the last years [31], taking into account different applications, sensors and data analysis methods. Techniques that combine measurements from different sensor systems can be found for water status assessment [32], yield forecasting [33–35] or plant monitoring [36,37]. Crop disease detection has also benefited from data fusion [38], but much of the research is based on feature fusion from single device data. In recent years, some deep learning architectures have been published for the fusion of two or more different sensors [39–41]. Nevertheless, as far as we know, very little research has been conducted on plant disease detection even though it is greatly needed. This would be particularly desirable with regard to the fusion of two very different types of sensors, such as MCF and hyperspectral imaging.

Melon crops (*Cucumis melo*), a valuable fruit worldwide, suffer from soft rot caused by the necrotrophic fungus *Botrytis cinerea*, which affects not only the fruit but also the plant during vegetative growth [42]. This pathogen is responsible for significant crop yield losses under favourable conditions, particularly in humid environments [43]. Traditional detection procedures, such as visual recognition by experts or biochemical methods, are sample- and time-consuming. Therefore, it is of interest to develop innovative, fast and reliable mechanisms for automated detection of this disease not only from a single sensor, but from several devices, taking advantage of the high availability of detection and data fusion methodologies. Thus, in this work, different deep learning data fusion architectures from two different sensors (MCF and hyperspectral reflectance) were used for the pre-symptomatically detection of soft rot infection in melon plants. The working hypotheses were: i) the models trained with the combination of diverse inputs are able to classify melon leaf samples as healthy or infected with high accuracy; and ii) these fusion models have a significantly better performance than single-sensor models. Accordingly, a set of architectures were designed for single-input and fusion modelling to compare their performance for this disease detection task, and their performance metrics were analyzed statistically for comparisons between architecture designs. The obtained results supported both hypothesis. Consequently, the integration of sensor fusion with a deep learning architecture significantly enhanced early soft rot detection in melon, presenting a feasible and effective solution for real-time, non-destructive diagnosis within the context of precision agriculture. To our best knowledge, this is the first time this procedure has been applied to melon plants infected with *B. cinerea*.

2. Materials and methods

2.1. Plant and fungal growth and infection

Melon plants (*Cucumis melo* L.) var. Rochet Panal (Semillas Barceló, Barcelona, Spain) were grown as previously described by Barón et al. [44]. *B. cinerea* Pers.: Fr strain B05.10 was grown in PDA (potato dextrose agar) plates (Biolife, Milan, Italy) at 25°C in light during 15 days. To prepare the inoculum, fungal conidia were collected from PDA

plates using 5 mL sterile distilled water. To remove remaining hyphae, the suspension was filtered through a 40 µm cell strainer (Corning, Glen-dale, AZ, USA). The concentration of the conidial suspension was adjusted to 10⁵ conidia per mL in half-strength filtered (0.45 µm; Merck Millipore, Darmstadt, Germany) grape juice (100% pure organic) using a Neubauer-improved counting chamber (Marienfeld Superior, Lauda-Königshofen, Germany). Four experiments were conducted, and a total of 24–28 plants were grown per trial. At the third week of growth, half of the plants were infected with *B. cinerea*, while the other half was considered as control, with no infection. Inoculation was carried out by placing four drops of 10 µL each of conidial suspension on a single leaf per plant, two on each side of the midrib. The control plants were mock-inoculated by placing drops of grape juice at the same spots as the infected plants. Each of the inoculated (or mock-inoculated) spots were considered as a sample for sensor data acquisition and deep learning modeling.

2.2. Sensor data collection

To monitor the evolution of the infection on the leaves, sensor data acquisition was carried out at 1 and 2 days post-inoculation (dpi), when no visible symptoms could be detected. The imagers used were the MCF device Open FluorCam FC 800-O (Photon Systems Instruments, Brno, Czechia) and the hyperspectral camera Pika L (Resonon, Bozeman, MT, USA). MCF image acquisition was carried out as described by Pérez-Bueno et al. [45], by placing the target leaf (each with the four spots) at a distance of 25 cm from the camera lens. Imaging of UV (355 nm) illuminated samples was set up for four wavelengths—440, 520, 680 and 740 nm—each one obtained from the average of nine frames of 800 ms exposure. Therefore, for each leaf, four 740 × 480 pixel resolution black and white images were acquired, one per wavelength. FluorCam v. 7.1.0.3 software (Photon Systems Instruments) operated this device.

Hyperspectral imaging was performed on the camera's spectral range—from 400 to 1000 nm, with a spectral resolution of 2.1 nm—in a push-broom configuration with a distance to the leaf again of 25 cm, homogeneously illuminated with four calibrated xenon lamps. Prior to the spectral measurement, a white reference (*W*) and dark current (*D*) were acquired to transform the raw data (*S*) to reflectance values (*R*):

$$R_{\lambda} = \frac{S_{\lambda} - D_{\lambda}}{W_{\lambda} - D_{\lambda}} \quad (1)$$

where S_{λ} is the light intensity from the sample at wavelength λ . Each hyperspectral image was a cube of spatial dimensions 900 × 857, each pixel with a spectrum of 300 variables.

The final, complete dataset comprised 739 samples (each sample corresponds to a *B. cinerea* inoculated—or mock-inoculated—spot), a shrank number from the original, intended grown plants (4 trials × 28 plants × 4 spots × 2 dpis = 896), as some samples were not in adequate conditions for image acquisition. From the 739 samples, 431 were of positive class (infected), and 308 of negative class (healthy). As illustrated by Fig. 1, each sample was composed of data from the two source devices:

- Four 64 × 64 MCF sub-images (patches), one for each wavelength, taken by extracting the patch around the inoculated spot. These dimensions were chosen as they fully capture the lesion development area without including potentially unaffected leaf tissue.
- One spectrum extracted from the average of a region of interest (average over 300–350 spectra) from the annulus shape just around the inoculated spot, but avoiding the juice grape drop, spanning the zone where early biochemical changes induced by *B. cinerea* infection are expected. The shape of this annulus varied per sampled spot.

These spatial dimensions for feature extraction were determined by the experimental inoculation protocol.

The structure of the collected data, in the form of mathematical tensors per sample for training a neural network, is as follows for each sensor:

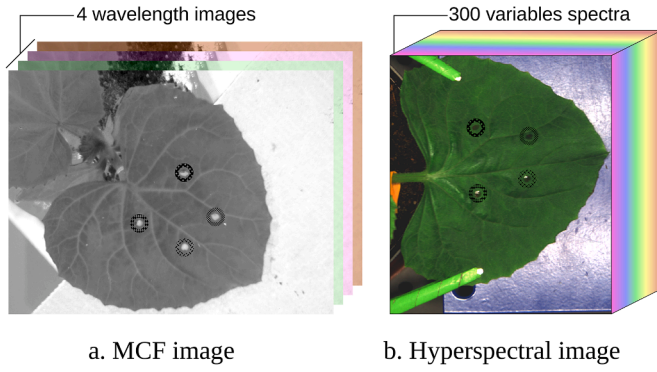


Fig. 1. Imaging of leaf samples using multicolor fluorescence (MCF) and hyperspectral cameras. The MCF image (a.) was comprised from images at four specific wavelengths: 440, 520, 680 and 740 nm. The hyperspectral cube contains reflectance spectra in the range of 400 to 1000 nm at a spectral resolution of 2.1 nm (*i.e.*, 300 variables). The marked regions of interest correspond to the same leaf spots in both images (inoculated or sham-inoculated spots). Each one of these spots is a sample in the final dataset.

1. **MCF camera.** As this device provides two-dimensional images on four different wavelengths, two data structures were tested:
 - A first-order tensor, $X_{\text{line}} \in \mathbb{R}^4$, representing the four wavelength values of the area around each infected spot, averaged from the manual extraction of a region of interest.
 - A third-order tensor, $X_{\text{img}} \in \mathbb{R}^{4 \times 64 \times 64}$, representing a 64×64 , four-channel image centered at each infected spot on the leaves. The channels in the image are created by stacking each individual image per wavelength.
2. **Hyperspectral camera.** A first-order tensor, $X_{\text{spectrum}} \in \mathbb{R}^{300}$, representing a discretized spectrum of 300 variables, equally distributing the sample reflectance values between 400 and 1000 nm.

2.3. Multi-sensor fusion architectures

Six different architectures were designed for the training of the classification models: 4 of them are base architectures, that always take data from a single sensor; while the other 2 are fusion architectures, that combine the feature extraction elements from the base architectures for higher complexity, thus using both hyperspectral and MCF data as input (Figs. 2 and 3). Briefly, these architectures are:

- **SPECTRALMLP:** A multilayer perceptron with a single hidden layer and fully connected, taking the reflectance spectrum as input.
- **SPECTRALCNN:** A deeper feature extraction and classification structure using one-dimensional convolution filters on the input reflectance spectrum.
- **FLUORMLP:** A multilayer perceptron with a single hidden layer and fully connected, taking the averaged MCF for the four wavelengths as input (refer to the end of Section 2.2).
- **FLUORCNN:** A convolutional neural network for images, taking as input the four-channel image of the infected spot comprised of the four wavelengths fluorescence data (refer to end of Section 2.2).
- **FUSIONNN_{LINE}:** A fusion architecture that uses the main structures from SPECTRALCNN and FLUORMLP.
- **FUSIONNN_{IMG}:** A fusion architecture built from SPECTRALCNN and FLUORCNN.

Each of these architectures is detailed in the following subsections. In all neural network function definitions, a trainable bias term is assumed for each weight tensor matrix, but it is omitted for readability.

2.3.1. SPECTRALMLP

This is a multilayer perceptron with all layers fully connected (Fig. 2a). The received input is the sample reflectance spectrum as a first

order tensor with 300 components (X_{spectrum}). Values are forwarded to a single hidden layer with 150 units—half the input, to set a value between input and output sizes [46], with a rectified linear unit (ReLU) activation function. The last layer is a single output neuron and a sigmoid function for the binary classification.

It is defined as $f_{\text{SpectralMLP}} : \mathbb{R}^{300} \rightarrow \mathbb{R}$:

$$f_{\text{SpectralMLP}}(X_{\text{spectrum}}) = \text{sigmoid} \left(\text{ReLU} \left(X_{\text{spectrum}}^T W_h \right) W_{out} \right), \quad (2)$$

with $W_h \in \mathbb{R}^{300 \times 150}$ as the single hidden layer, and $W_{out} \in \mathbb{R}^{150 \times 1}$ as the output layer.

2.3.2. SPECTRALCNN

A network that again processes spectral inputs but with a deeper architecture than that from SPECTRALMLP, using one-dimensional convolutions for feature extraction and fully connected layers for feature classification (Fig. 2b). The input is a sample reflectance spectrum X_{spectrum} , and it is forwarded to the feature extraction section, that comprises two one-dimensional convolution layers, followed by batch normalization, ReLU and a dropout with a probability value $p = .2$. The first convolution layer transforms the spectrum from one channel (as given in the input) to 30 channels, while the second convolution step downsizes the number of channels to 10. After this, the extracted features are flattened and forwarded to the classification section, with two fully connected layers until one single output neuron and a sigmoid function for binary classification.

This network is defined as $f_{\text{SpectralCNN}} : \mathbb{R}^{300} \rightarrow \mathbb{R}$:

$$f_{\text{SpectralCNN}}(X_{\text{spectrum}}) = \text{FC}_s \left(\text{flatten} \left(\text{FE}_s(X_{\text{spectrum}}) \right) \right), \quad (3)$$

Feature extraction is done by FE_s and classification is performed by FC_s . The module FE_s is defined as

$$\text{FE}_s(X_{\text{spectrum}}) = \text{BRD} \left(\text{BRD} \left(X_{\text{spectrum}}^T \otimes W_1 \right) \otimes W_2 \right), \quad (4)$$

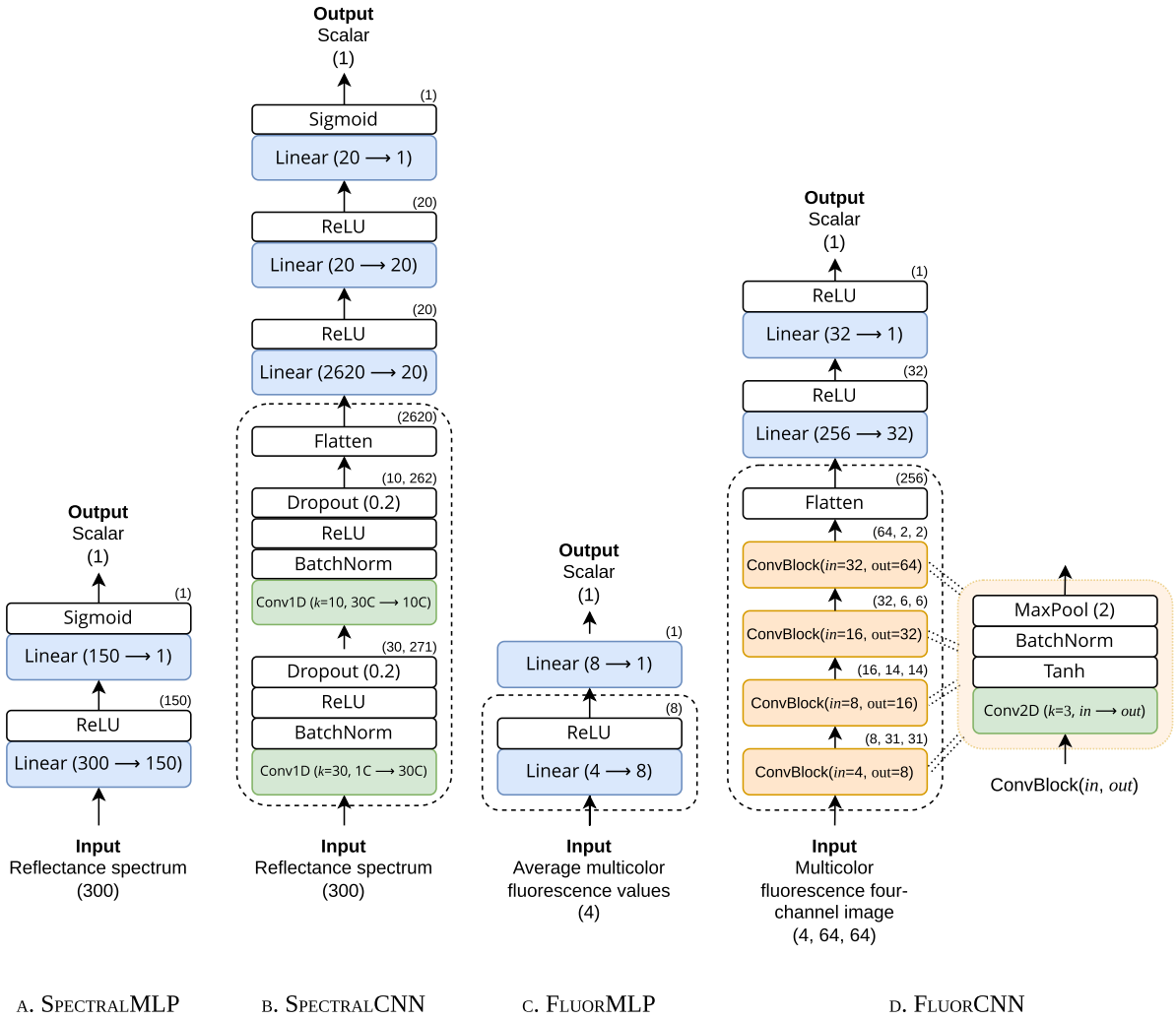


Fig. 2. Module diagrams of the base architectures. The networks using spectral input are SPECTRALMLP (a.) and SPECTRALCNN (b.); while FLUORMLP (c.) and FLUORCNN (d.) use MCF data as input. The modules around the dashed rectangles in b., c. and d. refer to feature extraction parts, and are the sections used in the fusion architectures in Fig. 3. All two-dimensional convolutional layers use square kernels, where k represents kernel dimensions $k \times k$. The numbers in parentheses above each module block represent the shape of its output tensor..

$$\text{BRD}(U) = \text{dropout}(\text{ReLU}(\text{batchnorm}(U))), \quad (5)$$

with $W_1 \in \mathbb{R}^{30 \times 2 \times 30}$ and $W_2 \in \mathbb{R}^{10 \times 2 \times 10}$ as two-dimensional convolution filters, with 30 and 10 channels, respectively. This feature extraction architecture (Eq. 4) is later used by both FUSIONNN_{LINE} and FUSIONNN_{IMG}.

Given a vectorized tensor $B \in \mathbb{R}^m$, the module FC_s is defined as

$$\text{FC}_s(B) = \text{ReLU}(\text{ReLU}(B^T W_1) W_2) W_{out}, \quad (6)$$

with $W_1 \in \mathbb{R}^{m \times 20}$, $W_2 \in \mathbb{R}^{20 \times 20}$, and $W_{out} \in \mathbb{R}^{20 \times 1}$ as the weights from the fully connected layers.

This spectral feature extraction architecture using one-dimensional convolutions is mainly based on previous works [47,48], also in spectral data fusion [40], for spectra from similar cameras and applications.

2.3.3. FLUORMLP

This is a very small non-linear model in the form of a multilayer perceptron (Fig. 3c) which receives an input of four wavelength-averaged values X_{line} and forward them into an 8-units hidden layer with ReLU activation. This hidden layer design is used as part of FUSIONNN_{LINE} (Section 2.3.5). It is defined as $f_{\text{FluorMLP}} : \mathbb{R}^4 \rightarrow \mathbb{R}$:

$$f_{\text{FluorMLP}}(X_{line}) = \text{sigmoid}(\text{FE}_{\text{fmlp}}(X_{line}) W_{out}), \quad (7)$$

$$\text{FE}_{\text{fmlp}}(X_{line}) = \text{ReLU}(X_{line}^T W_h) \quad (8)$$

with $W_h \in \mathbb{R}^{4 \times 8}$ as the single hidden layer, and $W_{out} \in \mathbb{R}^{8 \times 1}$ as the output layer.

2.3.4. FLUORCNN

As the MCF data can be shaped into four-channel images (refer to the end of Section 2.2), this architecture is based on common convolutional neural networks, with convolution and pooling layers (Fig. 2d). The input is the sample four-channel, 64×64 image (X_{img}) that is passed through two sections in the network: feature extraction and classification. Feature extraction contains four convolutional blocks with, in this order, two-dimensional convolution, hyperbolic tangent activation function, batch normalization and max pooling to half the block's input size. As with SPECTRALMLP, this feature extraction segment is also used by a fusion model, FUSIONNN_{IMG} in this case (Section 2.3.6).

FLUORCNN is defined as $f_{\text{FluorCNN}} : \mathbb{R}^{4 \times 64 \times 64} \rightarrow \mathbb{R}$:

$$f_{\text{FluorCNN}}(X_{img}) = \text{FC}_{\text{fcnn}}(\text{flatten}(\text{FE}_{\text{fcnn}}(X_{img}))), \quad (9)$$

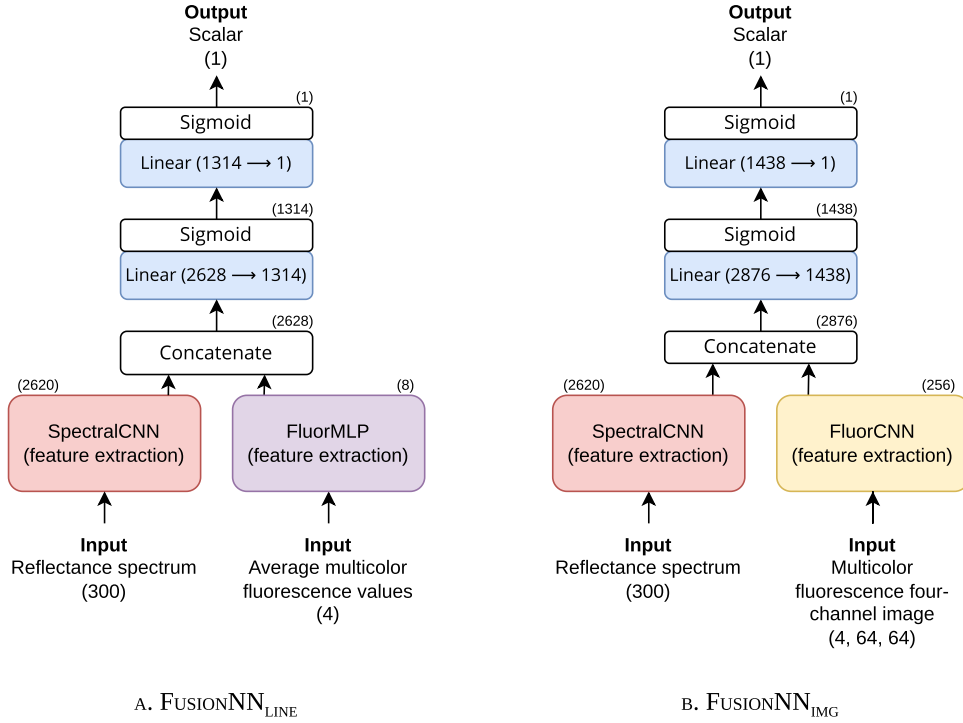


Fig. 3. Module diagrams of the fusion architectures, which use of both spectral and MCF data as input. While in both cases the spectral input is the full sample spectrum, MCF data are given as four wavelength-averaged values to FUSIONNN_{LINE} (a.), while data are structured as a four-channel image to FUSIONNN_{IMG} (b.). Feature extraction modules—SPECTRALCNN, FLUORMLP and FLUORCNN—refer to the modules in the dashed rectangles from the same networks in Fig. 2 (b., c. and d.). The numbers in parentheses above each module block represent the shape of its output tensor..

Feature extraction is done by FE_{fcnn} and classification is performed by FC_{fcnn} . The module FE_{fcnn} is defined as

$$FE_{fcnn}(X_{img}) = \text{ConvBlock}_{32 \rightarrow 64} \left(\text{ConvBlock}_{16 \rightarrow 32} \left(\text{ConvBlock}_{8 \rightarrow 16} \left(\text{ConvBlock}_{4 \rightarrow 8}(X_{img}) \right) \right) \right) \quad (10)$$

$$\text{ConvBlock}_{in \rightarrow out}(U) = \text{maxpool}_2(\text{batchnorm}(\tanh(U \otimes W_{in \rightarrow out}))) \quad (11)$$

with $U \in \mathbb{R}^{in \times h \times w}$ as the input tensor, and $W_{in \rightarrow out} \in \mathbb{R}^{out \times in \times 3 \times 3}$, as the convolution kernel.

Given a vectorized tensor $B \in \mathbb{R}^{256}$, the classification module FC_{fcnn} is defined as

$$FC_f(B) = \text{sigmoid} \left(\tanh(B^T W) W_{out} \right), \quad (12)$$

with $W \in \mathbb{R}^{256 \times 32}$ and $W_{out} \in \mathbb{R}^{32 \times 1}$ as the weights from the fully connected layers.

2.3.5. FUSIONNN_{LINE}

This fusion architecture takes two separate data inputs from the two sensors, perform feature extraction of them, and then model the classification of all the features into a binary output. Specifically, as shown in Fig. 3a., the spectrum (X_{spectrum}) and the four averaged MCF values (X_{line}) are the input for the two specific feature extraction modules, based on the SPECTRALCNN and FLUORMLP networks (the dashed rectangles from the same networks in Fig. 2b. and c.). The outputs of these modules—i.e., the extracted features—are then concatenated and forwarded to a fully connected layer, and then to the final single neuron output, for binary classification.

FUSIONNN_{LINE} is defined as $f_{\text{FUSIONNN}_{\text{LINE}}} : \mathbb{R}^{300} \times \mathbb{R}^4 \rightarrow \mathbb{R}$:

$$f_{\text{FUSIONNN}_{\text{LINE}}}(X_{\text{spectrum}}, X_{\text{line}}) = \text{sigmoid}(\text{ReLU}(FE_{\text{line}}(X_{\text{spectrum}}, X_{\text{line}})W)W_{out}) \quad (13)$$

$$FE_{\text{line}}(X_{\text{spectrum}}, X_{\text{line}}) = FE_s(X_{\text{spectrum}}) \oplus FE_{fmlp}(X_{\text{line}}) \quad (14)$$

with FE_{line} as the feature extraction function (making use of Eqs. 4 and 8), $\oplus : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n+m}$ as the operator one-dimensional tensor concatenation, $W \in \mathbb{R}^{2628 \times 1314}$ as a fully connected layer, and $W_{out} \in \mathbb{R}^{1314 \times 1}$ as the output layer.

2.3.6. FusionNN_{img}

The architecture FUSIONNN_{IMG} follows the same principle than FUSIONNN_{LINE}, with spectral and MCF inputs. Nevertheless, the latter is not the wavelength-averaged values, but the whole 64×64 chlorophyll fluorescence sub-image (refer to Section 2.2). Spectral feature extraction is done again with the same module used by SPECTRALCNN (modules in the dashed rectangles from the same networks in Fig. 2b.), while for MCF the feature extraction module from FLUORCNN is used (Fig. 2d.).

Similar to the previous fusion architecture, FUSIONNN_{IMG} is defined as $f_{\text{FUSIONNN}_{\text{IMG}}} : \mathbb{R}^{300} \times \mathbb{R}^{4 \times 64 \times 64} \rightarrow \mathbb{R}$:

$$f_{\text{FUSIONNN}_{\text{LINE}}}(X_{\text{spectrum}}, X_{\text{img}}) = \text{sigmoid}(\text{ReLU}(FE_{\text{img}}(X_{\text{spectrum}}, X_{\text{img}})W)W_{out}) \quad (15)$$

$$FE_{\text{img}}(X_{\text{spectrum}}, X_{\text{img}}) = FE_s(X_{\text{spectrum}}) \oplus FE_{fcnn}(X_{\text{img}}) \quad (16)$$

with FE_{img} as the feature extraction function (making use of Eqs. 4 and 10), $W \in \mathbb{R}^{2876 \times 1438}$ as a fully connected layer, and $W_{out} \in \mathbb{R}^{1438 \times 1}$ as the output layer.

2.4. Model training and statistical analyses

Models training was performed by neural network optimization, using the Adam algorithm [49] to minimize the binary cross-entropy loss function:

$$\mathcal{L}(y, \hat{y}) = -\frac{1}{N} \sum_{i=1}^N \left(y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right) \quad (17)$$

Table 1
Explored hyperparameters and their values.

Parameter	Values
Batch Size ($ B $)	{32, 64}
Learning Rate (η)	$\{10^{-5}, 10^{-4}, 10^{-3}\}$
Beta Pair (β_1, β_2)	$\{(0.9, 0.999), (0.95, 0.999)\}$
Epsilon (ϵ)	$\{10^{-8}, 10^{-7}\}$
Weight Decay (λ')	$\{0, 10^{-5}\}$

Table 2

Summary statistics of accuracy values. Accuracies correspond to the median of the top five local maxima from the validation accuracy curve of each model.

N	Min	Max	Mean	Median	Std. Dev.
4320	0.3716	0.8986	0.7206	0.7568	0.1024

with N as the batch size; y_i as the true binary label for sample i ; $\hat{y}_i$ the predicted probability label for sample i ; and $\log$ the natural logarithm.

To properly test the performance of the six architectures under several contexts, multiple models were trained rather than a single one by varying some training hyperparameters. The explored parameters are batch size, learning rate, beta pairs, epsilon and weight decay (L2 penalty). The explored values are shown in Table 1.

This set of architectures and hyperparameter choices results in 48 unique combinations. To account for the added variability arising from several stochastic processes within model training (initialization, data shuffling, etc.), three rounds of five-fold stratified cross-validation were conducted, resulting in fifteen repetitions per combination (*i.e.*, per trained model). All available samples were pooled and randomly assigned to training and validation folds at the individual sample level (*i.e.*, from all experiments and leaves) during cross-validation. Therefore, the total number of trained models and results to perform analyses on are $n = 6$ architectures $\times$ 48 parameter combinations $\times$ 15 repetitions = 4320. The total training time was approximately 170 hours, on a Intel® Core™ i7-13700K with a NVIDIA® GeForce RTX™ 4090 with 24 GB of vRAM, using Python 3.12.4 and PyTorch 2.4.0.

Each trained model was optimized for 300 epochs, as after supervised exploration this number was considered to provide the best balance between accuracy, computational efficiency and model complexity. For statistical analyses, the median of the top five local maxima from the validation accuracy curve was used as the performance metric for each model. The statistical analyses carried out on the results were Kruskal-Wallis test [50] as implemented in SciPy 1.15.2 [51]; and Dunn's test [52] with Bonferroni correction, implemented in the scikit-posthocs 0.11.3 library [53].

All models and training code are available on GitHub¹, while the data needed for training is openly available in Gutiérrez et al. [54].

3. Results

3.1. Statistical overview

The statistics for all the 4320 validation accuracy values from the trained models are summarized in Table 2. These yielded binary classification accuracies up to 0.8986, with a median of 0.7568 and a standard deviation of 0.1024.

The application of statistical tests was not direct, as previous normality checks were required. Histogram and Quartile-Quartile plots were created with all the data (Fig. 4).

The histogram (Fig. 4a) clearly shows a non-uniform distribution of the data, with two peaks at around the 0.60 and 0.80 marks. From

Table 3

Results of the Kruskal-Wallis tests for each parameter at $p < .05$. N.S.: no significant differences.

Parameter	p-value
Batch Size	0.2808 (N.S.)
Learning Rate	< 0.001
Beta Pairs	0.7639 (N.S.)
Epsilon	0.9463 (N.S.)
Weight Decay	0.7058 (N.S.)
Architecture	< 0.001

Table 4

Non-parametric ranking based on Dunn's test for the different values of learning rate and architecture. Different uppercase letters indicate significant differences within the parameter.

Parameter	Value	Mean accuracy
Learning rate (η)	10^{-3}	0.7838 (A)
	10^{-4}	0.7703 (B)
	10^{-5}	0.6871 (C)
Architecture	FUSIONNN _{LINE}	0.8068 (A)
	SPECTRALCNN	0.7932 (B)
	FUSIONNN _{IMG}	0.7838 (B)
	SPECTRALMLP	0.7500 (D)
	FLUORCNN	0.6486 (E)
	FLUORMLP	0.5831 (F)

the Quantile-Quantile plot (Fig. 4b), the deviations of the data from the theoretical line suggested that the values do not follow a normal distribution. This was further confirmed with the Anderson-Darling test for normality [55], implemented in SciPy 1.15.2 [51]. The result showed significant deviations from normality at the 5% significance level. This ruled out the null hypothesis that the data is normally distributed, so non-parametric tests were carried out.

3.2. Group analyses

For each hyperparameter and architecture, the Kruskal-Wallis non-parametric test was conducted. These parameters are the same mentioned in Section 2.4 (Table 1), plus the architecture. These results are summarized in Table 3.

No significant differences were found for the different values of batch size, beta pairs, epsilon and weight decay, so the values used for these parameters do not have significant influence on the model accuracy. Conversely, very strong differences ($p < .001$) resulted from the learning rate and the architecture used, therefore the specific learning rate value and, especially, the architecture selected from the six ones used have a significant effect on the exhibited performance. To better examine the group differences between the values of the parameters, Dunn's test [52] with Bonferroni correction was conducted for these two parameters. The results are gathered in Table 4.

The three tested values of the learning rate had a direct relation to the performance obtained, so higher learning rate (η) led to higher accuracy metrics, with significant differences between the three values (Table 4). For this parameter, the p -value was extremely low (Table 3), almost computationally collapsed to 0, and it was suspected that this behavior came from the large median accuracy difference between $\eta = 10^{-5}$ and the other values (0.6871 versus 0.7703 and 0.7838). We repeated the tests dropping all the samples for $\eta = 10^{-5}$, and a less extreme p -value was obtained ($p = .003890$), that still implied significant differences between $\eta \in \{10^{-3}, 10^{-4}\}$ as $p < .05$.

The results for the test on the architectures used (Table 4) showed that there were strong differences on median accuracy depending on the neural network selected for training, with a very large gap between the FLUOR architectures and the other ones (0.6486 to 0.7500). As with the learning rate η , we also suspected that this high difference would

¹ <https://github.com/suphan-barzani/fusion-architectures-disease-detection-melon.git>

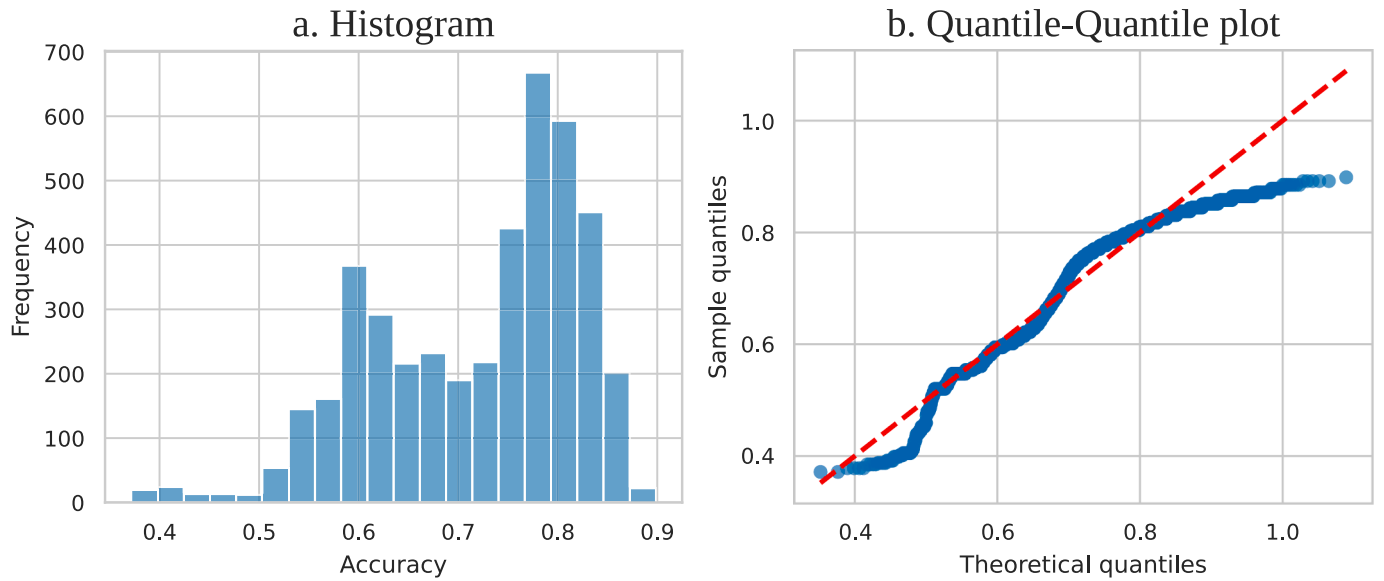


Fig. 4. Histogram (a.) and Quantile-Quantile (b.) plots of model accuracies. The dashed red line represents the reference line—i.e., where the theoretical quantiles should match the sample quantiles if the data were normally distributed. The points represent the sample quantiles, showing the actual accuracies from the models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5

Models with the highest performance metrics per architecture, along with the hyperparameters used in each one. Row order is consistent with that of Table 4. Reported accuracies correspond to the median of the top five local maxima from the validation accuracy curve.

Architecture	Batch Size	Learning Rate	Beta Pair	Epsilon	Weight Decay	Accuracy
FUSIONNN _{LINE}	32	10^{-4}	(0.95, 0.999)	10^{-8}	0	0.8919
SPECTRALCNN	64	10^{-4}	(0.95, 0.999)	10^{-7}	0	0.8919
FUSIONNN _{IMG}	64	10^{-3}	(0.95, 0.999)	10^{-8}	0	0.8986
SPECTRALMLP	32	10^{-3}	(0.95, 0.999)	10^{-8}	10^{-5}	0.8446
FLUORCNN	64	10^{-4}	(0.95, 0.999)	10^{-7}	10^{-5}	0.8446
FLUORMLP	32	10^{-3}	(0.90, 0.999)	10^{-8}	10^{-5}	0.6622

be the cause of the extremely low p -value obtained (Table 3), so the statistical tests were again repeated filtering out all samples from the FLUOR architectures, so only analyzing the performance of the other models. Nevertheless, very low p -values were repeatedly obtained as we removed from analyses the samples from the lower ranked architecture (letter grouping was always consistent with Table 4). The extreme case was a test with only two groups, FUSIONNN_{LINE} and SPECTRALCNN, and string significant differences were still found ($p < .001$). Therefore, the full analysis results in Table 4 were considered valid, as they remained consistent even after filtering out the extreme cases. Consequently, FUSIONNN_{LINE} was the highest-performing architecture (median accuracy of 0.8068), followed by the pair SPECTRALCNN and FUSIONNN_{IMG}. The two networks with the worst behavior were FLUORCNN and FLUORMLP, by a large margin, as mentioned, with accuracy values below the 0.65 mark.

3.3. Architecture analysis

From the complete set of results from all the trained models (4320 models, Section 3.1), the trainings with the highest accuracies *per architecture* were selected for further analysis. These are presented in Table 5.

All models achieved accuracy values over 0.84, except for FluorMLP, which yielded a maximum median validation accuracy of 0.6622, clearly below the rest. Attending to the parameters, for the three best architectures (from Table 4), FUSIONNN_{LINE}, SPECTRALCNN and FUSIONNN_{IMG}, no weight decay was added, while the remaining ones did have a weight decay of 10^{-5} . The values for the other hyperparameters varied among the possibilities (Table 1), except the beta pairs, for which

most of the values were (0.90, 0.999). For these best models in Table 5, their classification metrics are gathered in Table 6.

Table 6 reflects almost the same tendencies from the statistical analyses in Table 4, in terms of higher to lower performance ranking. In these specific cases, both fusion networks outperformed the single-sensor models. FUSIONNN_{LINE} reached the highest precision (0.9419) and specificity (0.9107) with a strong F1-score (0.9101), while FUSIONNN_{IMG} delivered the best F1-score (0.9153) and highest Kappa coefficient (0.7892), together with a competitive ROC-AUC (0.8936). Architectures using only MCF data, FLUORMLP and FLUORCNN, still stood out by a large margin as the worst performant data sources, with notably lower F1-scores and, in the case of FLUORMLP, a particularly poor ROC-AUC (0.5172) and Kappa (0.04163). Attending to the trade-off between sensitivity and specificity, FLUORCNN and FLUORMLP exhibited extreme sensitivity values (up to 1.0000) at the expense of specificity, which suggests a tendency to overprediction that does not translate into superior overall classification quality. This behavior indicates that these models are adopting a strategy of classifying most instances as positive to avoid missing true positives, thereby generating a high number of false positives. In the case of FLUORMLP, the ROC-AUC near random performance (0.5172) and extremely low Kappa (0.04163) confirm that the model has no real discriminatory capacity between classes, despite the perfect sensitivity.

Further training behavior can be analyzed by observing the evolution of loss and accuracy over the epochs, as presented in Fig. 5, for both training and validation statistics.

Fig. 5 shows close behavior patterns already observed in the scores from Table 6. Still, further information can be extracted from the

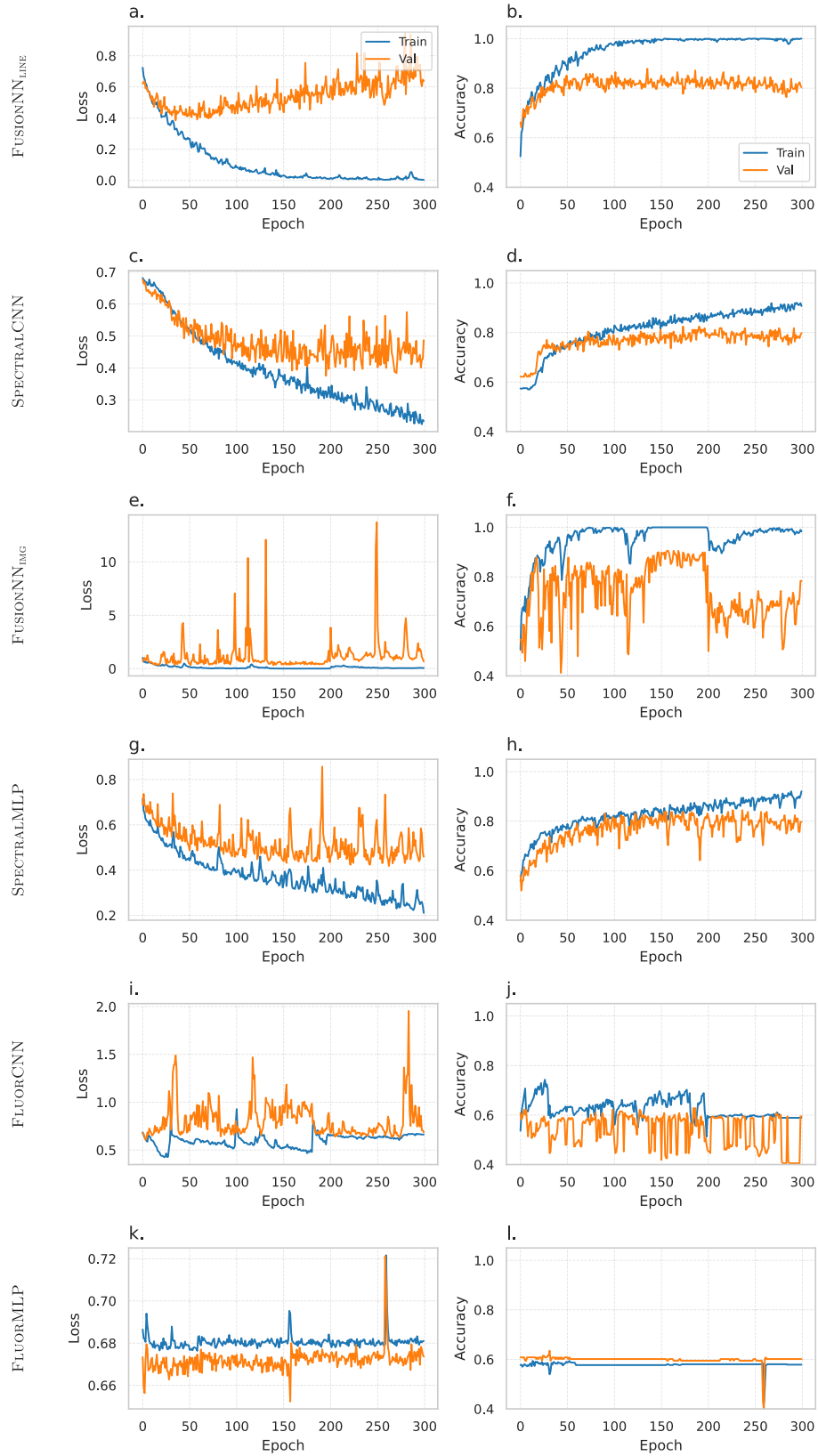


Fig. 5. Training and validation loss (left) and accuracy (right) over the configured 300 epochs for the best-performing model of each architecture: FUSIONNN_{LINE} (a–b), SPECTRALCNN (c–d), FUSIONNN_{IMG} (e–f), SPECTRALMLP (g–h), FLUORCNN (i–j), and FLUORMLP (k–l). Row order is consistent with that of Table 4.

Table 6

Classification metrics for the best model of each architecture. Row order is consistent with that of Table 4. Metrics are computed from the epoch with the highest validation accuracy.

Model	Epoch	Sensitivity	Specificity	Kappa	Precision	F1-Score	ROC-AUC
FUSIONNN _{LINE}	153	0.8804	0.9107	0.7749	0.9419	0.9101	0.8955
SPECTRALCNN	298	0.9022	0.8393	0.7415	0.9022	0.9022	0.8707
FUSIONNN _{IMG}	64	0.9205	0.8667	0.7892	0.9101	0.9153	0.8936
SPECTRALMLP	123	0.8519	0.8358	0.6868	0.8625	0.8571	0.8438
FLUORCNN	59	0.9481	0.6901	0.6446	0.7684	0.8488	0.8191
FLUORMLP	107	1.0000	0.03448	0.04163	0.6164	0.7627	0.5172

plots, with special emphasis in the correct, stable convergence of FUSIONNN_{LINE}, SPECTRALCNN and SPECTRALMLP. In all these three cases, the network quickly converges, and no more than 100–150 are needed to obtain the virtually lowest error (consequently, the highest accuracy). The training of this specific FUSIONNN_{LINE} model (Fig. 5g and h) showed a consistent tendency in training loss and accuracy, but a rather unstable behavior on the validation set. The poor, unstable training evolution of the best MFC models—FLUORCNN and FLUORMLP—is a reflection of the low average performance of these architectures analyzed in Table 4.

4. Discussion

Overall, it is presented an innovative and powerful method of deep learning models (four base architectures and two fusion architectures) based on a multi-sensor approach (MCF and hyperspectral reflectance images) designed for the rapid detection of melon leaves infected by soft rot caused by *B. cinerea*. In recent years, MCFI and hyperspectral reflectance imaging have become essential sensors for early disease detection. Moreover, it should be noted that, even though the proposed models could be considered shallow networks, the mean accuracy they achieved was high and comparable to other previously reported, as it will be discussed below. Using algorithms that necessitate reduced computing power could confer a substantial advantage in the field of disease detection, given that this would result in a reduction of both time and energy requirements.

In detail, the observed results confirm the original hypotheses on *i*) combining sensors for the reliable detection of infected melon leaves; and *ii*) the fusion architectures to perform better than single-sensor models, this also supported by statistical analyses. The case *i*) is confirmed by the high maximum accuracy values from Table 5, especially for the cases of FUSIONNN_{LINE} (0.8919), SPECTRALCNN (0.8919) and SPECTRALMLP (0.8446), that exhibited stable and consistent evolution during training, as seen in Figs. 5a–b, c–d and g–h. The best model from the FUSIONNN_{IMG} architecture showed an erratic training behavior in the validation dataset (Fig. 5e) and, although it peaks as the highest reached validation accuracy, over 0.89 around epoch 160 (Fig. 5f), this specific model does not clearly reflect the suitability of FUSIONNN_{IMG}. We include a supplementary figure for the second best model from this architecture (Fig. S1), that does show a much more stable training in both plots, with accuracy peaks over 0.87. These results are comparable with those previously published for the detection of plants infected by *B. cinerea* using one single-sensor measurements and deep learning analysis. For example, the detection of infected strawberries using hyperspectral imaging and one-dimensional CNNs achieved a maximum accuracy of 0.9686 [56]. In the case of pepper, infected leaves were detected with an overall accuracy of 0.8742 using the algorithm YOLO and hyperspectral imaging [57]. Finally, models based on multispectral cameras and CNNs identified cucumber infected plants with an overall accuracy of 0.901 [58].

The non-parametric analysis identified the architecture as one determinant of classification accuracy, and the group analysis supports the hypothesis *ii*), as one fusion architecture (FUSIONNN_{LINE}) formed a significantly different group from the other alternatives, ranking as

the best network with a validation accuracy of 0.8068. Although feature fusion—especially with deep learning modelling—can increasingly be found in recent articles for plant monitoring in agriculture, very few comparable articles use pure multi-sensor approaches in the recent literature. Architectures for the modeling of multi-source data have been reported for rice disease diagnosis [59], confirming, as in our results, that the combined model surpasses the individual ones (overall accuracy of 0.9531 of the fusion method vs. the 0.9125 of the most accurate single-sensor model). Also, multispectral imaging and MCF has been combined for the successful identification of citrus Huanglongbing (overall accuracy of 0.965) [60], modeling the data with CNNs, similar to our proposals. However, to the best of our knowledge, no other study on data fusion specific to the detection of *B. cinerea* infection could be found.

Although the absolute differences in mean accuracy between the two fusion architectures, FUSIONNN_{LINE} and FUSIONNN_{IMG}, are not large (0.8068 vs 0.7828, Table 4), they are significant. This difference can be explained by the dimensionality of the input data in each case. The MCF input is provided to FUSIONNN_{LINE} with a previous, supervised manual feature selection step through spatial averaging: the four MCF channels are averaged over the region of interest (the annuli from Fig. 1), giving the model only four scalar values to work with. This reduces the input dimensionality drastically and, in practice, filters the relevant information before it reaches the network. In contrast, FUSIONNN_{IMG} receives the full 64 × 64 four-channel image (16,384 dimensions) and must learn which pixels are relevant and which are not by itself. This high input dimensionality makes more difficult for the network to learn, as it is forced to learn from too many features. FUSIONNN_{LINE} benefits from manual feature engineering, while FUSIONNN_{IMG} is left to extract features automatically from raw high-dimensional data. This aligns with findings in plant disease classification [61] and hyperspectral imaging [62], where models with reduced input dimensionality demonstrate better generalization when working with limited training data. Additionally, this significantly better performance from the LINE fusion architecture against IMG could be advantageous when developing a dual-sensor for in-lab or even in-field applications, as the MCF device would not be required to capture whole images from the target.

No other parameter, apart from learning rate and architecture, had significant impact on the models' classification accuracy (Table 4). The fact that the higher, commonly default value of $\eta = 10^{-3}$ provides a significantly larger performance in accuracy is in concordance with analyses found in the literature [49,63]. Although incidentally half of the highest models per architecture had a higher η value, as seen in Table 5, it is clear from that table that none of the best models was configured with the lowest tested η , hence the marginally lower mean accuracy of $\eta = 10^{-5}$ in Table 4 (0.6871 versus 0.7703 from $\eta = 10^{-4}$). The second significantly different parameter, the architecture, had a very clear impact, as already discussed. But even more than the fusion architectures, Table 4 highlights the extremely low classification performance of the FLUOR networks, *i.e.*, the inability of MCF data alone to be modeled with deep learning for soft rot detection in melon leaves. The worst architecture was FLUORMLP, with a mean accuracy of 0.5831 (Table 4). This poor performance is also confirmed by the metrics evolution in Figs. 5k and l. These particular results do not align with those of other studies

that used the same technology [18,60,64,65]. This could be explained by the particular conditions involving the controlled infection of *B. cinerea* in leaves. The inoculation of this pathogen requires the use of a nutrient solution for higher chances of infection, thus disallowing the direct sensor data acquisition in exactly the same inoculation spot. It may be that the MCF imager is not sensible enough when acquiring data in the surroundings of the infection spots (refer to Fig. 1) to drive deep learning models into high classification metrics, hence the poorer outcomes from this sensor's models. The hyperspectral data alone, on the contrary, proved to be enough to train high-performant models (SPECTRALCNN's group ranked the second best in Table 4). Still, the MCF used jointly with spectral data incorporated enough useful information to induce a particular fusion architecture, FUSIONNN_{LINE}, to achieve significantly better results than any other model, especially compared with single-sensor architectures.

Regarding the classification metrics from the best models, FUSIONNN_{LINE} reached the highest sensitivity (0.9022) and the highest F1-score (0.9022), i.e., it missed the fewest diseased plants while still keeping most decisions consistent [64]. It was surpassed by FUSIONNN_{IMG} in terms of specificity (0.8500), in exchange for a small loss in sensitivity, thus producing fewer false alarms. Both fusion models achieved a Cohen's κ around 0.74, commonly interpreted as substantial agreement since it indicates the proportion of correct classifications that are not due to chance [66]. For an early disease detection tool, these statistics suggest that sensor fusion brings a real performance gain. Concerning the single-sensor architectures, SPECTRALCNN kept high sensitivity (0.8804) but lost more specificity (0.7321), leading to a higher rate of false-positive errors. This is also supported by its κ value of 0.62—moderate agreement—so this architecture is more prone to misclassifying healthy samples as diseased. SPECTRALMLP was the most balanced of the single-sensor models, requiring longest training (226 epochs) and still stayed below the fusion alternatives. As already discussed, both FLUORCNN and FLUORMLP performed far worse attending to their classification metrics.

Finally, the present study used 739 samples (431 infected, 308 healthy) taken under controlled lab-conditions. Thus, further analysis is required to study how the models would perform from data taken in the field, including other crops and larger amounts of samples, varying temperatures and growth stages. Another area for improvement is the implementation of MCF measures in the field, as this is currently challenging, although it is not strictly necessary, as it has already been discussed. Additionally, exploring simpler fusion mechanisms or employing pruning strategies could make the models lighter and faster for faster diagnosis, which may be important if real-time applications are required.

5. Conclusion

This study tested if the fusion of hyperspectral and MCF imaging data can detect soft rot caused by *Botrytis cinerea* infection in melon leaves in early stages; and if this detection is more effective than using single-sensor models. Six deep-learning architectures were trained and compared: four single-sensor networks and two fusion networks. The results support both hypotheses.

The best fusion model (i.e., the significantly better-performing architecture, FUSIONNN_{LINE}) reached a mean cross-validated accuracy of 0.8068 and a peak of 0.8919, significantly higher than every single-sensor alternative ($p < 0.001$). Architectures based only on fluorescence images or averaged fluorescence values performed worse, confirming that fluorescence alone was not sufficient under the present conditions.

These findings indicate that combining reflectance spectra with four fluorescence channels gives a measurable benefit for detecting soft rot. Therefore, sensor fusion with a deep-learning architecture improved early disease detection in melon and offers a prac-

tical path toward real-time, non-destructive diagnosis in precision agriculture.

CRedit authorship contribution statement

Salvador Gutiérrez: Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization; **Matilde Barón:** Writing – original draft, Resources, Project administration, Funding acquisition, Conceptualization; **María Trinidad Moreno-Martín:** Investigation, Conceptualization; **Mónica Pineda:** Writing – review & editing, Writing – original draft, Supervision, Investigation, Conceptualization.

Data availability

Data is available via an online dataset URL. References in the manuscript.

Declaration of competing interest

All authors declare that they have identified any financial assistance provided by a third party for the work presented; that they have no pending patents; and that the work will only be shared by citing the source appropriately.

Acknowledgements

The authors thank Dr. M.D. Fernández-Ortuño (IHSM UMA-CSIC) for kindly providing *Botrytis cinerea* Pers.: Fr strain B05.10.

Grant PID2022-139733OB-I00 funded by MICIU/AEI/10.13039/501100011033 and by “ERDF A way of making Europe” by ERDF/EU. MICIU: Spanish Ministerio de Ciencia, Innovación y Universidades; AEI: Agencia Estatal de Investigación; ERDF: European Regional Development Fund from the European Union.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.atech.2026.101892](https://doi.org/10.1016/j.atech.2026.101892)

References

- [1] J. Tardaguila, M. Stoll, S. Gutiérrez, T. Proffitt, M.P. Diago, Smart Appl. Digital Technol. Viticult.: A Rev. 1 (2021) 100005. <https://doi.org/10.1016/j.atech.2021.100005>
- [2] H. Moreno, D. Andújar, Proximal sensing for geometric characterization of vines: a review of the latest advances, Comput. Electron. Agr. 210 (2023) 107901. <https://doi.org/10.1016/j.compag.2023.107901>
- [3] X. Wang, H. Zeng, X. Yang, J. Shu, Q. Wu, Y. Que, X. Yang, X. Yi, I. Khalil, A.Y. Zomaya, Remote sensing revolutionizing agriculture: toward a new frontier, 2025, p. 107691. <https://doi.org/10.1016/j.future.2024.107691>
- [4] P. Tripodi, D. Massa, A. Venezia, T. Cardi, Sensing technologies for precision phenotyping in vegetable crops: current status and future challenges, Agronomy 8 (2018) 57. <https://doi.org/10.3390/agronomy8040057>
- [5] Q. Cao, C. Zhao, Z. Xu, P. Jiang, H. Yang, X. Meng, G. Yang, Review of the application of in-situ sensing techniques to address the tea growth characteristics from leaf to field, Int. J. Agric. Biol. Eng. 17 (2024) 1–11. <https://doi.org/10.25165/j.ijabe.20241701.8395>
- [6] V. Saiz-Rubio, F. Rovira-Más, A. Cuenca-Cuenca, F. Alves, Robotics-based vineyard water potential monitoring at high resolution, Comput. Electron. Agr. 187 (2021) 106311. <https://doi.org/10.1016/j.compag.2021.106311>
- [7] M. Pineda, M.L. Pérez-Bueno, M. Barón, Novel vegetation indices to identify broccoli plants infected with *Xanthomonas campestris* pv. *campestris*, front. Plant Sci. 13 (2022) 790268. <https://doi.org/10.3389/fpls.2022.790268>
- [8] A. Szabó, J. Tamás, A. Nagy, Spect. Estim. Chlorophyll non-invasive assessment apple orchards 10 (2024) 1266. <https://doi.org/10.3390/horticulturae10121266>
- [9] A.-K. Mahlein, Plant disease detection by imaging sensors—parallels and specific demands for precision agriculture and plant phenotyping, Plant Dis. 100 (2016) 241–251. <https://doi.org/10.1094/pdis-03-15-0340-fe>
- [10] A. Mohammad-Razdari, D. Rousseau, A. Bakhshipour, S. Taylor, J. Poveda, H. Kiani, Recent advances in E-monitoring of plant diseases, Biosens. Bioelectron. 201 (2022) 113953. <https://doi.org/10.1016/j.bios.2021.113953>

- [11] I. Hernández, S. Gutiérrez, I. Barrio, R.I. niguez, J. Tardaguila, In-field disease symptom detection and localisation using explainable deep learning: use case for downy mildew in grapevine, *Comput. Electron. Agr.* 226 (2024) 109478. <https://doi.org/10.1016/j.compag.2024.109478>
- [12] J.G.A. Barbedo, A review on the combination of deep learning techniques with proximal hyperspectral images in agriculture, *Comput. Electron. Agr.* 210 (2023) 107920. <https://doi.org/10.1016/j.compag.2023.107920>
- [13] M. Barón, M. Pineda, M.L. Pérez-Bueno, Picturing pathogen infection in plants, *Z. Naturforsch. C* 71 (2016) 355–368. <https://doi.org/10.1515/znc-2016-0134>
- [14] S. Thomas, M.T. Kuska, D. Bohnenkamp, A. Brugger, E. Alisaac, M. Wahabzada, J. Behmann, A.-K. Mahlein, Benefits of hyperspectral imaging for plant disease detection and plant protection: a technical perspective, *J. Plant Dis. Protect* 125 (2018) 5–20. <https://doi.org/10.1007/s41348-017-0124-6>
- [15] L. Wan, H. Li, C. Li, A. Wang, Y. Yang, P. Wang, Hyperspectral sensing of plant diseases: principle and methods, *Agronomy* 12 (2022) 1451. <https://doi.org/10.3390/agronomy12061451>
- [16] J. Wieme, K. Mollazade, I. Malounas, M. Zude-Sasse, M. Zhao, A. Gowen, D. Argyropoulos, S. Fountas, J.V. Beek, Application of hyperspectral imaging systems and artificial intelligence for quality assessment of fruit, vegetables and mushrooms: a review, *Biosyst. Eng.* 222 (2022) 156–176. <https://doi.org/10.1016/j.biosystemseng.2022.07.013>
- [17] M.L. Pérez-Bueno, M. Pineda, E.D. az Casado, M. Barón, Spatial and temporal dynamics of primary and secondary metabolism in *Phaseolus vulgaris* challenged by *Pseudomonas syringae*, *Physiol. Plant.* 153 (2015) 161–174. <https://doi.org/10.1111/pp1.12237>
- [18] M.L. Pérez-Bueno, M. Pineda, F.M. Cabeza, M. Barón, Multicolor fluorescence imaging as a candidate for disease detection in plant phenotyping, *Front. Plant Sci.* 7 (2016) 1790. <https://doi.org/10.3389/fpls.2016.01790>
- [19] M. Pineda, M.L. Pérez-Bueno, V. Paredes, M. Barón, Use of multicolour fluorescence imaging for diagnosis of bacterial and fungal infection on zucchini by implementing machine learning, *Funct. Plant Biol.* 44 (2017) 563–572. <https://doi.org/10.1071/fp16164>
- [20] M. Pineda, M. Barón, Assessment of black rot in oilseed rape grown under climate change conditions using biochemical methods and computer vision, *Plants* 12 (2023) 1322. <https://doi.org/10.3390/plants12061322>
- [21] N. Gambhir, A. Paul, T. Qiu, D.B. Combs, S. Hosseinzadeh, A. Underhill, Y. Jiang, L.E. Cadle-Davidson, K.M. Gold, Non-destructive monitoring of foliar fungicide efficacy with hyperspectral sensing in grapevine, *Phytopathology* 114 (2024) 464–473. <https://doi.org/10.1094/phyto-02-23-0061-r>
- [22] A. Brugger, F.I. Yamati, A. Barreto, S. Paulus, P. Schramowski, K. Kersting, U. Steiner, S. Neugart, A.-K. Mahlein, Hyperspectral imaging in the uv range allows for differentiation of sugar beet diseases based on changes in secondary plant metabolites, *Phytopathology* 113 (2023) 44–54. <https://doi.org/10.1094/phyto-03-22-0086-r>
- [23] M.B. Almoujahed, A.K. Rangarajan, R.L. Whetton, D. Vincke, D. Eylenbosch, P. Vermeulen, A.M. Mouazen, Detection of fusarium head blight in wheat under field conditions using a hyperspectral camera and machine learning, *Comput. Electron. Agr.* 203 (2022) 107456. <https://doi.org/10.1016/j.compag.2022.107456>
- [24] X. Wang, W. Yang, Y. Yang, M. Huang, Q. Zhu, Identification of tomato bacterial wilt severity based on hyperspectral imaging technology and spectrum transformer network, *Ecol. Inform.* 78 (2023) 102353. <https://doi.org/10.1016/j.ecoinf.2023.102353>
- [25] M.L. Pérez-Bueno, M. Pineda, C. Vida, D.F.-O. no, J. Torés, A. De, F.M. Vicente, M. Cazorla, Barón, Detection of white root rot in avocado trees by remote sensing, *Plant Dis.* 103 (2019) 1119–1125. <https://doi.org/10.1094/PDIS-10-18-1778-RE>
- [26] A.R. Allu, S. Mesapam, Impact of remote sensing data fusion on agriculture applications: a review, *Eur. J. Agron.* 164 (2025) 127478. <https://doi.org/10.1016/j.eja.2024.127478>
- [27] Y. Lu, P. Li, P. Wang, T. Li, G. Li, A method of rice yield prediction based on the QRBILSTM-MHSA network and hyperspectral image, *Comput. Electron. Agr.* 239 (2025) 110884. <https://doi.org/10.1016/j.compag.2025.110884>
- [28] Y. Lu, H. Zhou, P. Wang, E. Wang, G. Li, T. Yu, IMobileTransformer: a fusion-based lightweight model for rice disease identification, *Eng. Appl. Artif. Intell.* 161 (2025) 112271. <https://doi.org/10.1016/j.engappai.2025.112271>
- [29] V. Bolón-Canedo, L. Morán-Fernández, B. Cancela, A. Alonso-Betanzos, A review of green artificial intelligence: Towards a more sustainable future 599 (2024) 128096. <https://doi.org/10.1016/j.neucom.2024.128096>
- [30] C.E. Tripp, J. Perr-Sauer, J. Gafur, A. Nag, A. Purkayastha, S. Zisman, E.A. Bensen, Measuring the energy consumption and efficiency of deep neural networks: An empirical analysis and design recommendations, Technical Report, arXiv preprint, 2024. <https://doi.org/10.48550/arXiv.2403.08151>
- [31] J.G.A. Barbedo, Data fusion in agriculture: resolving ambiguities and closing data gaps, *Sensors* 22 (2022) 2285. <https://doi.org/10.3390/s22062285>
- [32] M.H. Kamarudin, Z.H. Ismail, N.B. Saidi, Deep learning sensor fusion in plant water stress assessment: A comprehensive review, *Appl. Sci.* 11 (2021) 1403. <https://doi.org/10.3390/app11041403>
- [33] A.A. Tesfaye, B.G. Awoke, T.S. Sida, D.E. Osgood, Enhancing smallholder wheat yield prediction through sensor fusion and phenology with machine learning and deep learning methods, *Agriculture-London*, 12, 2022. <https://doi.org/10.3390/agriculture12091352>
- [34] S. Fei, M.A. Hassan, Y. Xiao, X. Su, Z. Chen, Q. Cheng, F. Duan, R. Chen, Y. Ma, UAV-Based multi-sensor data fusion and machine learning algorithm for yield prediction in wheat, *Precis. Agric.* 24 (2023) 187–212. <https://doi.org/10.1007/s11119-022-09938-8>
- [35] K. Amankulova, N. Farmonov, E. Abdelsamei, J. Szatmári, W. Khan, M. Zhuran, J. Rustamov, S. Akhmedov, M. Sarimsakov, L. Mucsi, A novel fusion method for soybean yield prediction using Sentinel-2 and PlanetScope imagery, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 17 (2024) 13694–13707. <https://doi.org/10.1109/jstars.2024.3402114>
- [36] J. Reji, R.R. Nidamanuri, Deep learning based fusion of LiDAR point cloud and multispectral imagery for crop classification sensitive to nitrogen level, 1, IEEE, 2023. <https://doi.org/10.1109/MIGARS57353.2023.10064497>
- [37] Y. Liu, C. Wei, S.-C. Yoon, X. Ni, W. Wang, Y. Liu, D. Wang, X. Wang, X. Guo, Development of multimodal fusion technology for tomato maturity assessment 24 (2024) 2467. <https://doi.org/10.3390/s24082467>
- [38] M. Ouhami, A. Hafiane, Y. Es-Saady, M.E. Hajji, R. Canals, Computer vision, IoT and data fusion for crop disease detection using machine learning: a survey and ongoing research, *Remote Sens.* 13 (2021) 2486. <https://doi.org/10.3390/rs13132486>
- [39] P. Mishra, D. Passos, Deep multiblock predictive modelling using parallel input convolutional neural networks, *Anal. Chim. Acta* 1163 (2021) 338520. <https://doi.org/10.1016/j.aca.2021.338520>
- [40] S. Gutiérrez, J. Fernández-Navales, T. Garde-Cerdán, S. Marín-San Román, J. Tardaguila, M.P. Diago, Multi-sensor spectral fusion to model grape composition using deep learning, *Inform. Fusion* 99 (2023) 101865. <https://doi.org/10.1016/j.inffus.2023.101865>
- [41] K. Sriprateep, S. Khonjun, R. Pitakaso, T. Srichok, S. Sala-Ngam, Y. Sriethep, S. Gonwirat, P. Luesak, S. Matitopanam, C. Chueadee, et al., Hybrid adaptive multiple intelligence system (HybridAMIS) for classifying cannabis leaf diseases using deep learning ensembles, *Smart Agric. Technol.* 9 (2024) 100535. <https://doi.org/10.1016/j.atech.2024.100535>
- [42] K. Bi, Y. Liang, T. Mengiste, A. Sharon, Killing softly: A roadmap of *Botrytis cinerea* pathogenicity, *Trends Plant. Sci.* 28 (2023) 211–222. <https://doi.org/10.1016/j.tplants.2022.08.024>
- [43] M.D.C. Orozco-Mosqueda, A. Kumar, A.E. Fadji, O.O. Babalola, G. Puopolo, G. Santoyo, Agroecological management of the grey mould fungus *Botrytis cinerea* by plant growth-promoting bacteria, *Plants* 12 (2023) 637. <https://doi.org/10.3390/plants12030637>
- [44] M. Barón, M.T. M.-M. í, M. Pineda, Uncovering *Botrytis cinerea*-induced physiological changes in melon plants using multi-sensor imaging approaches, *Plant Stress*, 2025. <https://doi.org/10.1016/j.jstress.2025.100769>
- [45] M.L. Pérez-Bueno, E. Granum, M. Pineda, V. Flors, P. Rodríguez-Palenzuela, E. López-Solanilla, M. Barón, Temporal and spatial resolution of activated plant defense responses in leaves of *Nicotiana benthamiana* infected with *Dickeya dadantii*, *Front. Plant Sci.* 6 (2016) 1209. <https://doi.org/10.3389/fpls.2015.01209>
- [46] K.G. Sheela, S.N. Deepa, Review on methods to fix number of hidden neurons in neural networks, 2013, p. 425740. <https://doi.org/10.1155/2013/425740>
- [47] S. Gutiérrez, A. Wendel, J. Underwood, Ground based hyperspectral imaging for extensive mango yield estimation, *Comput. Electron. Agr.* 157 (2019) 126–135. <https://doi.org/10.1016/j.compag.2018.12.041>
- [48] S. Gutiérrez, A. Wendel, J. Underwood, Spectral filter design based on in-field hyperspectral imaging and machine learning for mango ripeness estimation, *Comput. Electron. Agr.* 164 (2019). <https://doi.org/10.1016/j.compag.2019.104890>
- [49] D.P. Kingma, Adam: A method for stochastic optimization, Technical Report, arXiv preprint, 2014. <https://doi.org/10.48550/arXiv.1412.6980>
- [50] W.H. Kruskal, W.A. Wallis, Use of ranks in one-criterion variance analysis, *J. Am Stat. Assoc.* 47 (1952) 583–621. <https://doi.org/10.2307/2280779>
- [51] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al., Scipy 1.0: Fundamental algorithms for scientific computing in python, *Nat. Methods* 17 (2020) 261–272. <https://doi.org/10.1029/s2199-1006.1.sor-life.a7056644.v1.rysreg>
- [52] O.J. Dunn, Multiple comparisons using rank sums, *Technometrics* 6 (1964) 241–252. <https://doi.org/10.2307/1266041>
- [53] M.A. Terpilowski, Scikit-posthocs: pairwise multiple comparison tests in Python, *J. Open Source Softw.* 4 (2019) 1169. <https://doi.org/10.21105/joss.01169>
- [54] S. Gutiérrez, M. Pineda, M. Barón, M.T. Moreno-Martín, Dataset for the deep learning fusion architectures for disease detection in melon, 2024. <https://doi.org/10.5281/zenodo.15669670>
- [55] T.W. Anderson, D.A. Darling, Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes, 1952, pp. 193–212. <https://doi.org/10.1214/aoms/1177729437>
- [56] S.-W. Chun, D.-J. Song, K.-H. Lee, M.-J. Kim, M.S. Kim, K.-S. Kim, C. Mo, Deep learning algorithm development for early detection of *Botrytis cinerea* infected strawberry fruit using hyperspectral fluorescence imaging, *Postharvest Biol. Tec.* 214 (2024) 112918. <https://doi.org/10.1016/j.postharvbio.2024.112918>
- [57] D. Kapetas, E. Kalogeropoulou, P. Christakakis, C. Klaridopoulos, E.M. Pechlivani, Comparative Evaluation of AI-based multi-spectral imaging and PCR-based assays for early detection of *Botrytis cinerea* infection on pepper plants, *Agriculture-London*, 15, 2025. <https://doi.org/10.3390/agriculture15020164>
- [58] N. Giakoumoglou, E. Kalogeropoulou, C. Klaridopoulos, E.M. Pechlivani, P. Christakakis, E. Markellou, N. Frangakis, D. Tzovaras, Early detection of *Botrytis cinerea* symptoms using deep learning multi-spectral image segmentation, *Smart Agric. Technol.* 8 (2024) 100481. <https://doi.org/10.1016/j.atech.2024.100481>
- [59] R.R. Patil, S. Kumar, Rice-fusion: a multimodality data fusion framework for rice disease diagnosis, *IEEE Access* 10 (2022) 5207–5222. <https://doi.org/10.1109/access.2022.3140815>
- [60] C. He, X. Li, Y. Liu, B. Yang, Z. Wu, S. Tan, D. Ye, H. Weng, Combining multi-color fluorescence imaging with multispectral reflectance imaging for rapid citrus Huanglongbing detection based on lightweight convolutional neural network using a handheld device, *Comput. Electron. Agr.* 194 (2022) 106808. <https://doi.org/10.1016/j.compag.2022.106808>

- [61] S. Quan, J. Wang, Z. Jia, M. Yang, Q. Xu, Ms-net: A novel lightweight and precise model for plant disease identification, *Front. Plant Sci.* 14 (2023) 1276728. <https://doi.org/10.3389/fpls.2023.1276728>
- [62] K.-H. Liu, M.-H. Yang, S.-T. Huang, C. Lin, Plant species classification based on hyperspectral imaging via a lightweight convolutional neural network model, *Front. Plant Sci.* 13 (2022) 855660. <https://doi.org/10.3389/fpls.2022.855660>
- [63] S.J. Reddi, S. Kale, S. Kumar, On the convergence of Adam and beyond, Technical Report, arXiv preprint, 2019. <https://doi.org/10.48550/arXiv.1904.09237>
- [64] S. Sankaran, R. Ehsani, Detection of huanglongbing-infected citrus leaves using statistical models with a fluorescence sensor, *Appl. Spectrosc* 67 (2013) 463–469. <https://doi.org/10.1366/12-06790>
- [65] C.B. Wetterich, R.F. D.O. Neves, J. Belasque, R. Ehsani, L.G. Marcassa, Detection of huanglongbing in Florida using fluorescence imaging spectroscopy and machine-learning methods, *Appl. Opt.* 56 (2016) 15–23. <https://doi.org/10.1364/ao.56.000015>
- [66] J. Carletta, Assessing agreement on classification tasks: The kappa statistic, Technical Report, arXiv preprint, 1996. <https://doi.org/10.48550/arXiv.cmp-lg/9602004>