Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Data availability

Links listed below are the data deposition URLs from Data availability section. Please verify the links are valid. This page will not appear in the article PDF file or print. They are displayed in the proof pdf for review purpose only.

https://github.com/DumitruScutelnic/MME Dataset Link: https://github.com/DumitruScutelnic/MME

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Highlights

Multi-model ensembles for object detection in multispectral images: A case study for precision agriculture

Computers and Electronics in Agriculture xxx (xxxx) xxx

Dumitru Scutelnic*, Claudia Daffara, Riccardo Muradore, Martin Weinmann, Boris Jutzi

- Scalable MME pipeline for object detection in multispectral imagery.
- Adaptive integration of spectral bands without dimensionality reduction.
- Comparative analysis of multispectral strategies, including false color.
- Validation on real and public agricultural data for disease detection.
- MME applicable in the range UV, VIS, NIR, SWIR, and LWIR for reliable analysis.

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

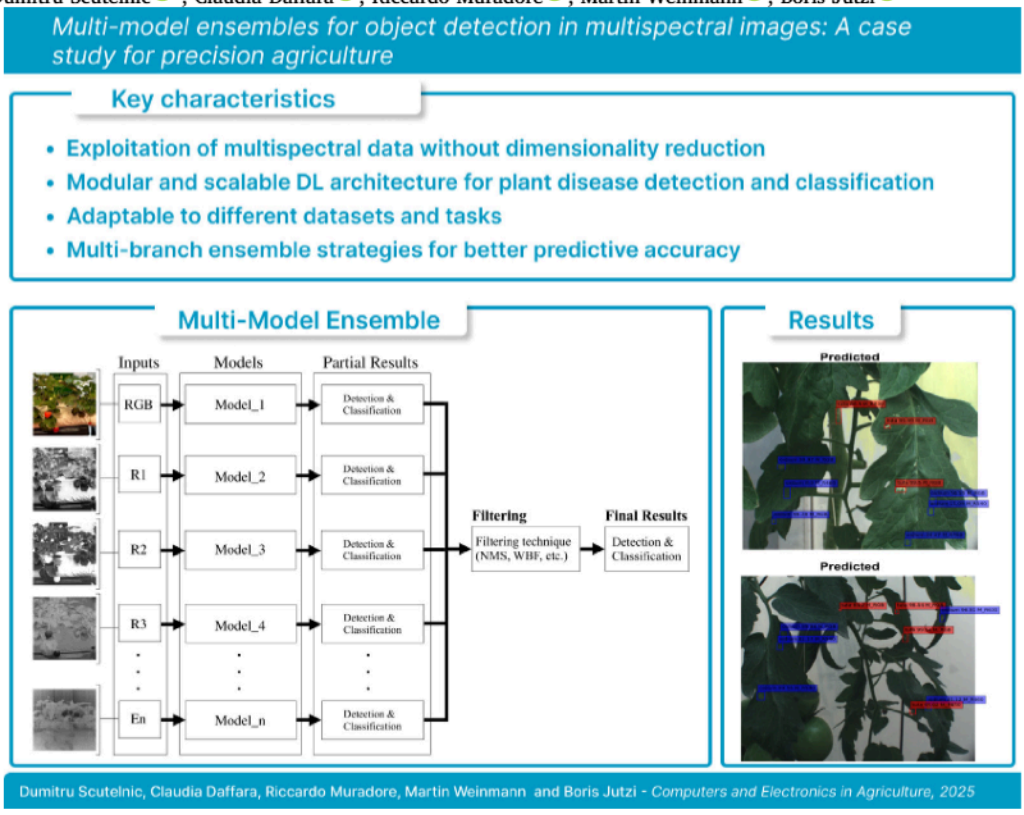
journal homepage: www.elsevier.com/locate/compag

Graphical Abstract

Multi-model ensembles for object detection in multispectral images: A case study for precision agriculture

Computers and Electronics in Agriculture xxx (xxxx) xxx

Dumitru Scutelnic*, Claudia Daffara, Riccardo Muradore, Martin Weinmann, Boris Jutzi



Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Multi-model ensembles for object detection in multispectral images: A case study for precision agriculture

Dumitru Scutelnic [a],*, Claudia Daffara [a], Riccardo Muradore [b], Martin Weinmann [c], Boris Jutzi [c]

[a] Department of Computer Science, University of Verona, Verona, Italy
[b] Department of Engineering for Innovation Medicine, University of Verona, Verona, Italy
[c] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology, Karlsruhe, Germany

ARTICLE INFO

Dataset link: https://github.com/DumitruScutelnic/MME

Keywords:
Deep Learning
Multi-Model Ensemble
Precision agriculture
Multispectral dataset
Plant disease detection

ABSTRACT

Every year, 20%–40% of the global harvest is lost to pests and diseases, underlining the need for rapid and accurate diagnosis. Precision agriculture exploits intelligent devices, such as robots and drones, to enable early detection of pathogens through non-destructive imaging techniques and AI processing. In this study, we exploit Deep Learning techniques for handling multispectral images in agriculture field. In particular, we introduce an adaptive Multi-Model Ensemble framework that processes multispectral data without dimensionality reduction, fully exploiting spectral information to improve early disease detection. Furthermore, several comparisons with dimensionality reduction and data combinations were conducted, exploring different image stack configurations to find the optimal solution in disease detection. We validated our approach on a dataset of tomato plants affected by *Tuta Absoluta* and *Leveillula Taurica*, where it improves the ability of disease identification and classification even at early developmental stages, offering promising perspectives for phytosanitary monitoring and sustainable resource management.

1. Introduction

The Food and Agriculture Organisation of the United Nations estimates that 20% to 40% of the global harvest is lost to pests and diseases each year, despite the application of about two million tons of pesticides (Manida, 2022). The prevailing method of detecting pests and diseases generally depends mainly on the knowledge and experience of farmers; the most common strategy employed to date for risk management is using plant protection chemicals Brand et al. (2009). Unfortunately, if pathogens and pests are at an advanced stage of progression in the crop, the effectiveness of pesticides and fungicides may be compromised and contribute to developing resistance by pathogens and pests (Lucas et al., 2015; Hawkins et al., 2019). For this reason, applications of products to counteract are often preventive and applied throughout the field, thus introducing an additional fixed cost of more than 15% of the operating cost, besides being a health and environmental hazard (Tudi et al., 2021). Since the incidence of plant diseases and pests is increasing, new, accurate, and timely methods must be employed to identify symptoms as early as possible.

In the literature, several works use non-destructive techniques such as multispectral imaging (Singh et al., 2020; De Silva and Brown, 2023b), hyperspectral imaging (Rayhana et al., 2023; Zhang et al., 2020), and thermal imaging (Xu et al., 2006; Pineda et al., 2021) for mapping pathologies in plants. These technologies have long been known in various fields of application: industrial (Daffara et al., 2020), medical (Lu and Fei, 2014), and cultural heritage (Daffara et al., 2021; de Manincor et al., 2020), and more recently in agri-tech (Tsouros et al., 2019). Multispectral and, especially, hyperspectral imaging techniques can generate vast amounts of data when applied in real-world scenarios, such as agriculture. For instance, it is possible to acquire thousands of images from a single hectare, where each sample generates a multidimensional data (stack of images); in addition, it must also be considered spatial and temporal information. In multispectral imaging, each sample is represented as a stack of images (x, y, n_λ), where x and y denote the spatial dimensions, and n_λ indicates the number of spectral bands acquired. Each band corresponds to a specific central wavelength, typically with a spectral bandwidth ranging from 50 nm to 100 nm. In contrast, hyperspectral imaging generates a three-dimensional data cube at narrow intervals, typically around 10 nm, providing a high spectral resolution across a broad wavelength range (Bhargava et al., 2024).

* Corresponding author.
E-mail address: dumitru.scutelnic@univr.it (D. Scutelnic).

https://doi.org/10.1016/j.compag.2025.111213
Received 20 April 2025; Received in revised form 20 October 2025; Accepted 6 November 2025

Without artificial intelligence (AI), processing such large volumes of data would take significantly longer, potentially extending to several weeks, resulting in delayed decision-making and an increased risk of suboptimal or ineffective treatments. Expertise in disease identification and classification is crucial, as high-performance imaging alone is insufficient without efficient data processing. Integrating automated processing systems based on AI becomes essential to handling large datasets within an acceptable time frame. The synergy between advanced optical techniques and AI enables the development of automated disease detection systems, facilitating real-time monitoring and precise mapping of diseases, pathogens, and biotic or abiotic stresses in both spatial and temporal domains. By enabling targeted and localised interventions, AI-driven methods help optimise resource use and minimise the excessive application of agrochemicals. Additionally, these technologies reduce reliance on manual labour and allow continuous operation in greenhouses and open fields through fixed or mobile platforms such as farm vehicles and unmanned ground vehicles (UGVs) (Halder et al., 2023; Musanase et al., 2023). This marks a pivotal phase in agriculture's digital transformation, where AI and advanced technologies empower farmers to make data-driven decisions. This paradigm, known as Precision Agriculture (PA) (Nowak, 2021), is fundamental to enhance productivity while reducing resource consumption and mitigating environmental and health impacts. Thereby, the task of disease identification and classification can be treated as an object detection task.

Among the traditional approaches for object detection, a straightforward solution consists of keypoint matching and subsequent geometric verification via a least-squares estimation of affine projection parameters (Lowe, 2004). Furthermore, the Histogram of Oriented Gradients (HOG) detector (Dalal and Triggs, 2005) has been proposed, which is based on the idea of evaluating local histograms of image gradient orientations on a dense grid. More advanced solutions accounting for more variability of considered objects are represented by the Deformable Part-Based Model (DPM) (Felzenszwalb et al., 2008) and Implicit Shape Model (ISM) (Leibe et al., 2006) approaches. With the advent of modern Deep Learning (DL) approaches and their capability of learning robust feature representations, significant advancements have also been presented for the task of object detection (Zou et al., 2023). In this regard, two-stage detectors focus on first identifying regions of interest (ROI) where objects might be present and subsequently classifying the objects within these regions as well as refining their locations. The Region-based Convolutional Neural Network (R-CNN) (Girshick et al., 2014) uses a selective search algorithm to propose ROI in an image and then a Convolutional Neural Network (CNN) to classify each ROI. Fast R-CNN (Girshick, 2015) improves on R-CNNs by using a shared CNN to process the entire image and all ROIs in one step with multi-task learning of a classifier and a bounding box regressor rather than processing each ROI independently. This significantly reduces the computation time required for object detection. Faster R-CNN (Ren et al., 2015) is the modified version of Fast R-CNN. Instead of the selective search used by the Fast R-CNN to derive ROIs, the Faster R-CNN relies on a CNN-based approach represented by a Region Proposal Network (RPN). The RPN uses anchor boxes for object detection; the generation of regions is faster and better adapted to the data compared to the selective search used in R-CNN and Fast R-CNN via an external algorithm. RPN takes image feature maps as input and generates a series of object proposals, each with an objectivity score as output. It also enables scale-invariant detections using Feature Pyramid Networks (FPNs). The features are shared with the detection network to estimate the object's position, significantly accelerating detection. Two-stage detectors deliver accurate results, but typically at higher computational costs and with more sophisticated network architectures. In some applications, however, a trade-off between accuracy and inference time is required. Hence, one-stage detectors have been proposed to solve the task of object detection in one single step, addressing both the prediction of bounding box coordinates and the prediction of class probabilities for the derived bounding boxes. A prominent approach among the one-stage approaches is the You Only Look Once (YOLO) detector (Redmon et al., 2016) and its variants.

To date, DL has emerged as the best-performing method for image analysis, providing robust feature extraction and effectively addressing non-linear relationships in data and achieving state-of-the-art results in detection, segmentation, and classification tasks compared to traditional machine learning techniques (Kumar et al., 2024). It produces good results and has great potential for applications in many real-world fields, including the agricultural field (Bharman et al., 2022) for yield prediction, weed detection, plant stress (Gao et al., 2020), and disease detection (Saleem et al., 2019). Nowadays, most of the approaches in literature identify plant diseases using only RGB images (Abade et al., 2021).

For early detection of diseases in plants, until today, there are no commercial products, and research is facing significant challenges:

- The state of evolution of the pathology, its irregular shape and spread, and different patterns of the lesion for the same type of infection or pathology, making feature learning a difficult task.
- The size of infection in the early stages is in the millimetre scale, requiring imaging at high spatial resolution.
- Relevant spectral responses are not only in the visible range, requiring multispectral imaging.
- If the plant is large, in a whole-plant shot some regions may be out of focus, leaves will cover each other and may be facing in all directions. Moreover, background is typically heterogeneous and non-uniform due to the complex environments (greenhouse and open-field), making the acquisition of a quality dataset challenging and time-consuming.

In addition, early symptoms appear in limited plant areas, only on some leaves or fruits, and it is challenging to obtain samples of infected plants, as special conditions are necessary to prevent the spread of infection and cross-contamination. Furthermore, the involvement of expert personnel is essential to ensure accurate annotation and diagnosis of the pathological signs.

Current public datasets for plant disease analysis primarily focus on classification tasks, as summarised in Table 1. Most of these datasets are based on RGB images, thus limiting spectral information and generalisation under variable real-world conditions. Only a few datasets provide multispectral or hyperspectral data, allowing for more detailed physiological and biochemical assessments of plants. Furthermore, many datasets are acquired in controlled or semi-controlled environments, which may restrict their applicability in open-field scenarios. Finally, annotation types range from simple classification labels to bounding boxes and pixel-level segmentations.

Up to now, it is unclear which is the best approach to handle multidimensional data, such as multispectral or hyperspectral data, via standard DL and transfer learning methods, as these typically only accept input images in the 1- or 3-channel format, for which they were designed and trained (Vali et al., 2020). The literature suggests using PCA (Pesaresi et al., 2024), which reduces the dimensionality of the data by projecting them into an orthogonal feature space and selecting there a few components that explain most of the variance. This approach can work well when the essential features are mainly found in the three principal components, but there is the risk of losing relevant information.

Multispectral imaging systems for agricultural applications are increasingly equipped with more sensors to capture reflectance (multispectral imaging) and emissivity (i.e. thermal imaging), capable of performing multiple tasks simultaneously, including assessment of plant health, crop maturity, through to pathogen detection and identification of various diseases (Scutelnic et al., 2024a). In these cases, spectral responses can have distinctive characteristics in different bands with

Table 1
Summary of major plant disease datasets with image types, annotations, and key details.

Dataset	Image type	No. of images	Spectrum/Bands	Annotations	Crops/Main diseases	Key notes
PlantVillage (2015) (Hughes and Salathé, 2015)	RGB	50000 images	Visible (RGB)	Classification (14 crops, 26 diseases)	Various crops, 26 diseases	Controlled environment images
PlantDoc (2019) (Singh et al., 2019)	RGB	2600 images	Visible (RGB)	Classification, bounding boxes	13 species, 17 diseases	"In-the-wild" images with real backgrounds
Raza et al. (2015)	Thermal + Stereo RGB	71 plants	Thermal + Visible	Healthy vs infected, temporal classes	Tomato (Oidium neolycopersici)	14-day monitoring in controlled environment
Wang et al. (2019)	RGB	286 original + augmented	Visible (RGB)	Bounding boxes, segmentation masks	Tomato (various diseases)	Natural, complex backgrounds
VddNet (2020) (Kerkech et al., 2020)	Multispectral + Depth Map	Not specified	Visible (RGB), NIR + Depth	Pixel-level segmentation (4 classes)	Grapevine (vineyard diseases)	UAV multispectral + depth, orthophoto alignment, parallel encoder architecture
Khan et al. (2021)	Hyperspectral VIS–NIR	Not specified	400–1000 nm	Quantitative disease severity	Wheat (powdery mildew)	Expert validation, greenhouse, early classification
Fernández et al. (2021)	Multispectral	20 multispectral images	Blue, Green, Red, Red-Edge, NIR	Pixel-level healthy vs infected	Cucumber (downy mildew)	Greenhouse, SVM classifier
Potato Stress Dataset (2021) (Butte et al., 2021)	UAV Multispectral	360 patches	RGB + Green, Red, Red-Edge, NIR	Bounding boxes (healthy vs stressed)	Potato	High-resolution images in open-field
Barros et al. (2022)	RGB + Multispectral + Thermal	Not specified	5 multispectral bands + thermal	Pixel-level vineyard row segmentation	Grapevine	Multimodal UAV dataset in open-field
Peng et al. (2022)	Multispectral (A-MSI handheld device)	Not specified	400–1000 nm (multispectral)	Healthy vs infected classification	Cassava – Cassava brown streak disease (CBSD)	Early detection 28 days post-inoculation using spatial–spectral ML handheld device, field-adapted
Vélez et al. (2023)	UAV Multispectral	16500 images	Blue, Green, Red, Red-Edge, NIR	Georeferenced, infection clusters	Grapevine (Botrytis cinerea)	Spatial mapping in open-field
De Silva and Brown (2023a)	Multispectral	2600 images	VIS + NIR (K590, K665, K720, K850)	Healthy vs stressed classification	Various crops	Open-field data
Giakoumoglou et al. (2023)	RGB (real-field and greenhouse)	659	Visible (RGB)	Bounding boxes (Tuta absoluta)	Tomato (Tuta absoluta)	Ensemble detection with Faster R-CNN/RetinaNet and improved mAP via WBF (from 0.58 to 0.70)
Georgantopoulos et al. (2023)	5 Multispectral + RGB	263 multispectral stack	RGB + multispectral	Bounding boxes	Tomato (Tuta absoluta, Leveillula taurica)	Disease detection in greenhouse
Ryckewaert et al. (2023)	Hyperspectral VIS–NIR	205 images	400–900 nm	Chemical measures and health annotations	Grapevine	Spectral-physiological modelling in controlled environment
PlantSeg (2024) (Wei et al., 2024b)	RGB	11400 images	Visible (RGB)	Pixel-level disease symptom segmentation	Various crops	In open-field dense segmentation
PlantWild (2024) (Wei et al., 2024a)	RGB	Not specified	Visible (RGB)	Segmenta-tion/classification	Various crops	In open-field images with complex backgrounds
Zhang et al. (2024)	Hyperspectral VIS–NIR + SWIR	Daily images for 7 days	400–1000 nm + 900–1700 nm	Infection stages, pathogen load	Tomato (Xanthomonas perforans)	Early disease detection (within 2h)
Estrada et al. (2025)	Multispectral + Spectral Reflectance	Not specified	5 multispectral bands + 350–2500 nm reflectance	Biometric measurements, calibrated data	Avocado, Olive, Grapevine	Leaf dehydration monitoring, multimodal data
Saccuti et al. (2025)	Multispectral	172 images	6 multispectral bands	Healthy vs diseased classification	Grapevine (Lambrusco varieties)	Field conditions, Flavescence dorée and Esca

different physics meanings. For this reason, merging these data with heterogeneous information content can lead to learning difficulties if a single AI model is used. Consequently, it is necessary to modify the backbone part of the model architecture (1) to make it possible to accept image data with more than 3-channels as input by changing the first convolution layer (i.e. 2D convolution or Conv2D), or (2) to compute a maximum of 3 spectral bands at a time.

1.1. Related works

Recent studies have explored the use of deep neural networks for multispectral image processing, highlighting the potential in complex tasks such as object detection and segmentation in multi-channel contexts. Sa et al. (2016) introduced DeepFruits, an automatic fruit detection system based on DL and transfer learning, designed to operate

on RGB and NIR images. The authors explore two multimodal fusion strategies: early fusion, which combines data at the input level before feature extraction, and late fusion, which combines predictions derived from separate networks for each modality. Using a pre-trained CNN adapted via transfer learning demonstrates that multimodal integration beyond RGB improves the performance in complex environments. Liu et al. (2016) proposed a framework based on deep CNN for pedestrian detection by integrating visible and thermal images, exploiting fusion at different stages of the architecture (early, halfway, and late fusion). Almahasneh et al. (2022) developed a multi-layer multi-task CNN architecture for multispectral images capable of performing detection and segmentation, even in the presence of weakly supervised annotations. Kulkarni's pioneering work (2023) (Kulkarni, 2023) demonstrated the effectiveness of CNNs even in traditional pixel-wise classification. Each pixel is treated as a spectral vector that is transformed into a representation used by a CNN for classification to obtain accurate feature maps. However, the study notes a limitation in the quantity of features that can be processed due to the finite size of the transformed image. A taxonomy of DL-based fusion strategies for object detection using multimodal sensors can be found in Liu et al. (2024). Input fusion (combination of multimodal images at pixel-level, before feature extraction) is simple and fast, but sensitive to noise and requires precise alignment. Halfway fusion combines intermediate features from multispectral images and is sub-grouped into: Early halfway fusion (at the input level through channel concatenation), which is fast but sensitive to noise; Late halfway fusion (combination of features extracted from different networks) that balances robustness and generalisation, but is more complex; and Hybrid halfway fusion (at the intermediate level of the network, after partial feature extraction) offers a good balance between efficiency and semantic depth, but can be difficult to optimise. Decision-level fusion (combination of final results predicted by separate models) is flexible, but less computationally efficient.

The integration of heterogeneous information from RGB, thermal, and multispectral sensors is challenging due to temporal and spatial alignment, feature compatibility and computational constraints. To overcome these limitations, this work introduces a scalable Multi-Model Ensemble (MME) framework, based on a late or decision-level fusion strategy, specifically designed to fully exploit multispectral variability for reliable and early plant disease detection.

1.2. Aims of this work

This research focuses on the implementation and comparison of state-of-the-art DL methods for extracting information from multidimensional multispectral data, with the aim of exploiting the entire available spectrum and maximising performance in plant disease detection. To this end, approaches involving dimensionality reduction through the selection and combination of spectral bands are analysed, as well as methods that operate directly on the entire multidimensionality of the data. The comparative analysis of the results obtained made it possible to identify the most effective strategies for the optimal use of spectral information to improve the predictive accuracy of the models. To validate the proposed method in agriculture, we used, as case study, a dataset consisting of RGB images and multispectral stack of five images in monochrome format, acquired in narrow bands in the VIS-NIR on tomato plants affected by two distinct diseases.

In addition, we propose a scalable pipeline based on a Multi-Model Ensemble (MME) approach, which uses a decision-level fusion strategy and allows the entire multidimensionality of multispectral data to be processed or, if necessary, only the most informative bands, in order to maximise detection performance. Several configurations were defined and tested, including:

- **Multi-channels input:** processing the entire multispectral stack or a subset (as 8-channel or 6-,5-, 4-channels input) by adapting the first network layer (Conv2D). This configuration evaluates the use of all available spectral information as input.
- **RGB input:** training models on RGB images with and without enhancement to explore the potentiality of RGB space colour and transfer learning approach.
- **Single spectral-band input:** training models on individual spectral bands images to identify the most informative for disease detection.
- **Independent mixing input:** combining the RGB and the multispectral images that showed the best performance in the previous phase as independent inputs to the same model. This configuration tests whether the most informative spectral components can be effectively integrated without full decision-level fusion.
- **Input-level fusion (early fusion):** combining channels into three-channel images or using vegetation indices to represent the most informative spectral combinations in simplified form. This approach integrates data directly at the input level, prior to feature extraction, allowing the model to process false-colour composites and vegetation indices as a single input.
- **Decision-level fusion (late fusion):** MME method that combines the results of multiple specialised models, preserving the entire multispectral stack and integrating predictions to achieve robust and generalisable detection performance.

2. Materials and methods

For this research, we used the open-source dataset published by Georgantopoulos et al. (2023), which includes multispectral imaging samples of tomato plants acquired in a greenhouse. These samples are annotated with two infections, *Tuta Absoluta* (labelled as *Tuta*) and *Leveillula Taurica* (labelled as *Oidium*).

2.1. Description of infections

In the first weeks of development, the *Tuta Absoluta* midge's offspring are larvae that feed on the leaves and cause complete necrosis and consequent defoliation. This pathogen is estimated to lead to a 50% to 100% loss in tomato plant yield (Ghaderi et al., 2019). *Leveillula Taurica*, known as *powdery mildew* or *oidiopsis taurica*, usually infects only fully developed mature leaves. The main symptoms are yellow irregular spots (approximately 10 mm to 15 mm in diameter). It is generally located on the abaxial side of the leaves, and a thin, light brown mycelium forms on the adaxial surface due to the emergence of conidiophores through the stomata (Aegerter et al., 2015). Several studies report yield losses of up to 40% (Guzman-Plazola et al., 2003) in greenhouse tomatoes and 80% to 100% (Desneux et al., 2010) in field tomato crops. Therefore, control in the initial stages is essential to limit damage as much as possible.

2.2. Dataset

The considered dataset of Georgantopoulos et al. (2023) includes 263 multispectral image samples of tomato plants acquired in a greenhouse. The images have a size of 1776 × 2368 pixels with 8-bit dynamic range; they have been acquired with the MUSES9-MS-PL camera equipped with a CMOS sensor with sensitivity in the range from 360 nm to 1100 nm. Each sample consists of a stack comprising an RGB image and 5 narrow band spectral images in the VIS-NIR corresponding to reflectance information acquired in bands with centre wavelength and width, of 460 nm (30 nm), 540 nm (30 nm), 630 nm (40 nm), 850 nm (50 nm), and 980 nm (50 nm). An example of a stack is shown in Fig. 1. From now on, for the sake of simplicity, we refer to these bands as R460, R540, R630, R850 and R980, where R means reflectance; the RGB channels are indicated with R, G, B. A total of 229 samples were annotated, resulting in 1757 cases of *Tuta Absoluta* infections and only 239 cases of *Leveillula Taurica* infections, resulting in a class imbalance.

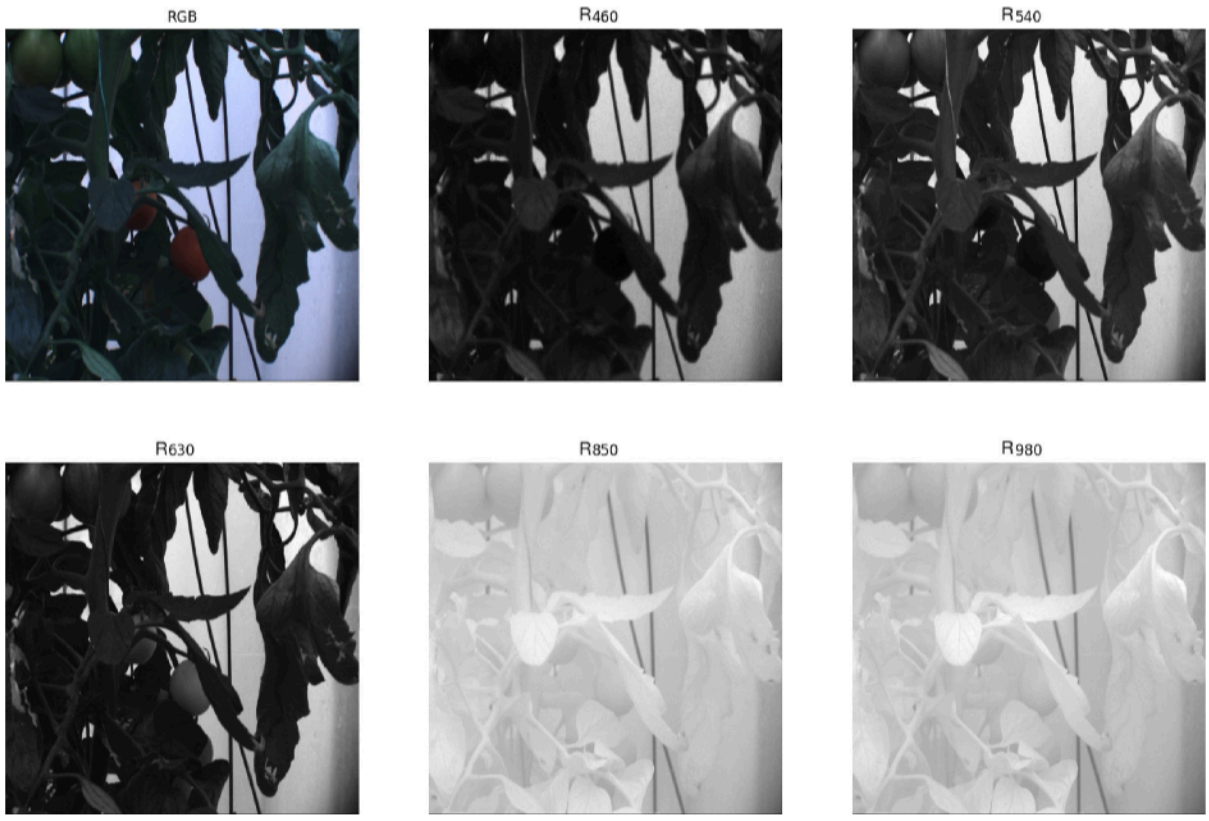


Fig. 1. Example of a sample stack from dataset by Georgantopoulos et al. (2023), including RGB and spectral images at different wavelengths (460 nm, 540 nm, 630 nm, 850 nm, and 980 nm) representing the in-band reflectance.

In case of representative and sufficient training data, improving images through enhancement techniques does not significantly improve the learning of the model and, therefore, the achieved accuracy (Santana et al., 2024; Salvi et al., 2021). Thus, high performance is usually only achieved if the data is acquired optimally. The considered dataset, however, has only a limited number of samples. The RGB images reveal that image acquisition was not performed in the optimal conditions, given the non-uniform and yellowish-grey background, leading to several false positives. Thus, we also employed techniques of image enhancement by taking into account that RGB colour representations are less robust with respect to changes in illumination. In particular, we focus on transformations of the RGB colour dataset to more advanced representations. Among different options (Weinmann and Weinmann, 2019), we select approaches based on colour constancy, since colour constancy is able to assign a particular and stable colour percept to an object, irrespective of its surroundings and illumination (Gegenfurtner et al., 2024). On the one hand, we use an approach relying on the GreyWorld assumption according to which the average reflectance of surfaces in the scene is considered achromatic (Buchsbaum, 1980). On the other hand, we use Edge-Based Colour Constancy (EBCC) relying on the hypothesis that the average edge difference in a scene is achromatic (van de Weijer et al., 2007). Both the GreyWorld and the EBCC approaches transform RGB colour images to a representation with three entries per pixel.

The dataset was split into 80% training samples and 20% test samples, so that all experiments rely on the same disjoint image samples for training and testing.

2.3. Methods and performed analyses

In this study, we adopt the Faster R-CNN architecture (Ren et al., 2015), a two-stage object detection framework that enables both localisation and classification of target instances within an image. We employed pre-trained ResNet-50 backbones to initialise the model weights and used the Adam optimiser with a fixed learning rate of 0.0001 across all experiments. To investigate the sensitivity of the models to training dynamics, we varied the number of training epochs (30, 50, and 100) and batch sizes (1, 5, 10, 15, and 20). This analysis aimed to evaluate whether performance metrics improved or degraded under different training configurations, particularly in scenarios where baseline results were already satisfactory.

2.3.1. Multi-channels method

The first method that we investigated is the adaptation of a standard Faster R-CNN architecture by modifying the first convolutional layer. Specifically, we modified the first convolutional layer (Conv2D) of the backbone to accommodate non-standard input formats, enabling the processing of images with an arbitrary number of channels, thus, allowing a single 8-channel image (Full Stack) to be passed to the model as a stack of RGB and VIS-NIR multispectral monochromatic images (R460, R540, R630, R850, and R980). The pipeline is shown in Fig. 2. We also considered image stacks of RGB images and subgroups of the spectral bands (S1, S2, S3 and S4), resulting in a 6, 5, and 4 channel image stack as follows:

- Full Stack: [R, G, B, R460, R540, R630, R850, R980]
- Stack S1: [R, G, B, R460, R540, R630]
- Stack S2: [R, G, B, R540, R630]
- Stack S3: [R, G, B, R540]
- Stack S4: [R, G, B, R630].

In this method, the network processes all channel jointly and learns spectral–spatial relationships across the entire input stack in a single forward pass.

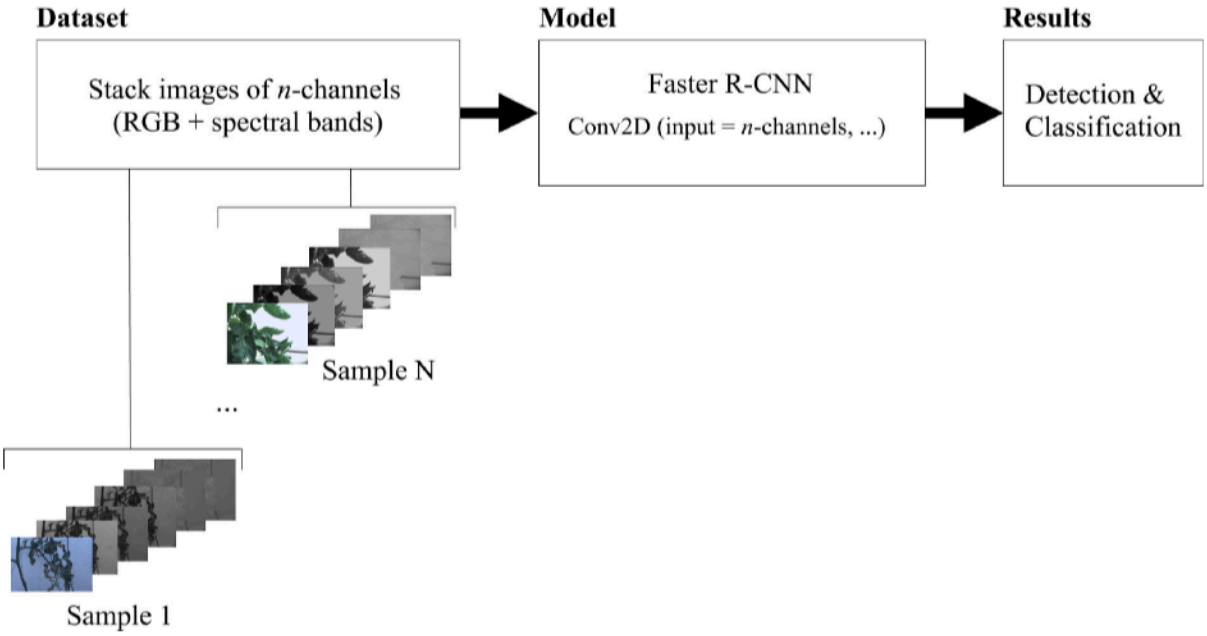


Fig. 2. Pipeline of the multi-channels method that handles image stacks made by *n*-channels (RGB plus multispectral images).

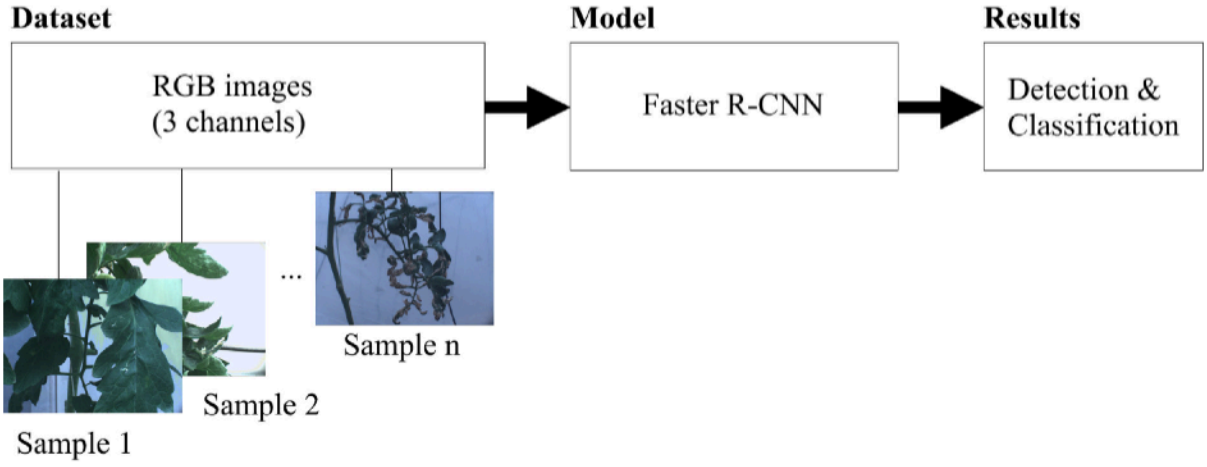


Fig. 3. Classic pipeline of the RGB method, where only RGB images are used as input.

2.3.2. RGB method

The subsequent analysis uses only the original, unprocessed RGB images. This reflects a classic and most commonly approach for detection and classification, where only 3-channel RGB images are used, as shown in Fig. 3.

In a second analysis, we applied image enhancement techniques to improve the quality since, as mentioned, the original RGB images were acquired under non-optimal conditions and calibration. Specifically, we employed the GreyWorld and EBCC methods to enhance colour balance and ensure consistency across all RGB images. This pre-processing phase allows us to assess the impact of image quality optimisation on model performance, particularly in scenarios such as an open field or greenhouse with different environmental conditions.

2.3.3. Spectral band method (single-channel)

In the third method, individual models were trained separately for the single spectral band (R460, R540, R630, R850 or R980), as illustrated in Fig. 4. The model learns features based solely on spatial and intensity variations within that specific band. This analysis enabled the evaluation of learning rates and the identification of the most relevant spectral bands for our purpose. The results indicate that the manifestation of a specific plant disease exhibits a distinct spectral signature within particular wavelength ranges. Consequently, not all spectral bands contribute equally to the classification task, highlighting the importance of selecting the most informative wavelengths to optimise model performance.

2.3.4. Independent mixing method

In the fourth method, we have used all the RGB images and all spectral images in the different bands R460, R540, R630, R850 and R980 as independent input. Fig. 5 shows the architecture of this pipeline.

We used as inputs the RGB images with the most relevant spectral images, demonstrating the highest performance to detect the two tomato diseases, which manifest their symptoms predominantly in the R540 and R630, during the learning process in previous stages, when considered individually in the spectral band method. RGB and multispectral bands are not stacked into a single input but are considered as separate inputs during training, Specifically, the dataset was constructed by generating independent inputs that are passed to the model based on the following combinations:

- D1: RGB images and all spectral images (R460, R540, R630, R850, and R980)

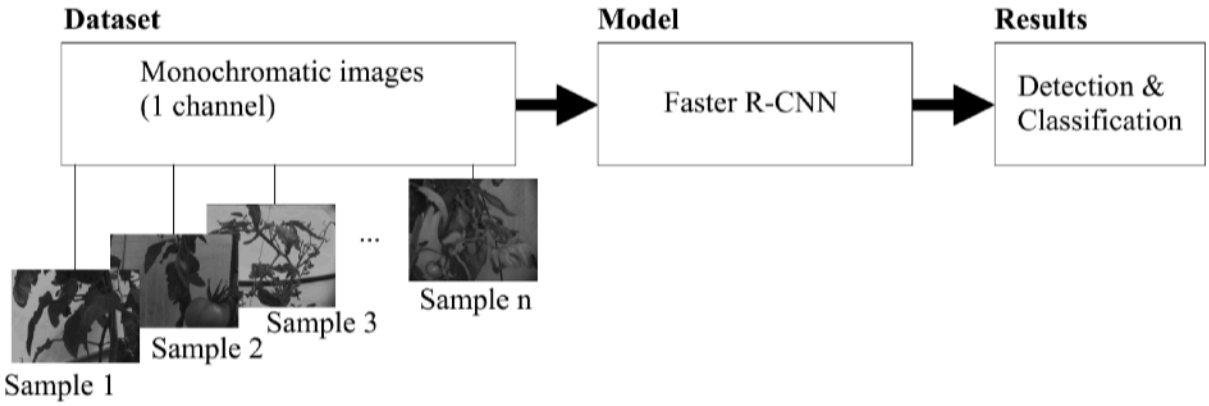


Fig. 4. Pipeline of the spectral band method, where images in a single spectral band are used as independent input.

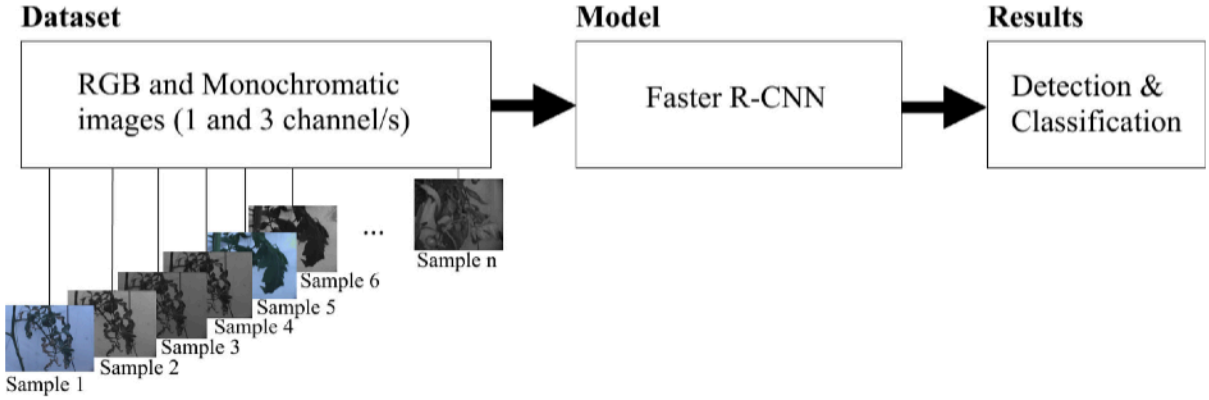


Fig. 5. Pipeline of the independent mixing method, where RGB and multispectral images are used together as independent input.

- D2: RGB images and spectral images (R460, R540, and R630)
- D3: RGB images and spectral images (R540 and R630)
- D4: RGB images and R540 spectral images
- D5: RGB images and R630 spectral images.

2.3.5. *Reconstructed 3-channel method*

In the fifth method, we combined the spectral images with RGB images, removing one or more channels from the RGB images to obtain false-colour images. In addition, we combined the bands by taking 3 at a time to reconstruct 3-channel images. This selection was guided by the bands that yielded the best results in previous analyses, leading to the following combinations:

- FC1: [R630, R, G]
- FC2: [R540, R, G]
- FC3: [R850, R, G]
- FC4: [R, R540, R630]
- FC5: [G, R540, R630]
- FC6: [R630, R540, R460].

Fig. 6 presents a schematic representation of the processing pipeline, where C1, C2, and C3 correspond to different combinations of three channels. Each channel can represent a specific multispectral band or a channel from the RGB image.

This method differs significantly from training the model on single-band (monochromatic) inputs. When the three bands are combined into a single input, the network can also learn inter-band relationships, i.e., learn how the spectral contrast between the bands correlates with disease symptoms. This interplay between bands may be key for detecting spectral cues that may not be apparent in a single band.

In addition, we also computed the Normalised Difference Vegetation Index (NDVI) and included it as an alternative input representation. NDVI is a well-established vegetation index widely used in remote sensing to assess plant health and photosynthetic activity, as it emphasises the contrast between red (absorbed by chlorophyll) and NIR (reflected by healthy vegetation). In our context, NDVI was computed using the R630 (Red) and R850 (NIR) bands, as follows

$$NDVI = \frac{R850 - R630}{R850 + R630}. \quad (1)$$

It was integrated into the experimental framework to investigate the effectiveness in enhancing the detection and identification of diseased regions, as compared to the use of raw spectral input. This approach aims to explore the potential of vegetation indices as low-dimensional, interpretable inputs for disease identification tasks.

2.3.6. *The proposed multi-model ensemble method*

In the sixth method, we have developed a scalable Multi-Model Ensemble approach consisting of several independent models that were previously trained. This pipeline uses multispectral data integrated with visible information (RGB) and spectral data to make partial predictions independently of each other, considering the entire dataset or subgroups:

- M1: RGB, R460, R540, R630, R850, and R980
- M2: RGB, R540, and R630
- M3: RGB and R540
- M4: RGB and R630.

Subsequently, the results obtained from the different models are merged and selected according to the highest scores, thereby optimising

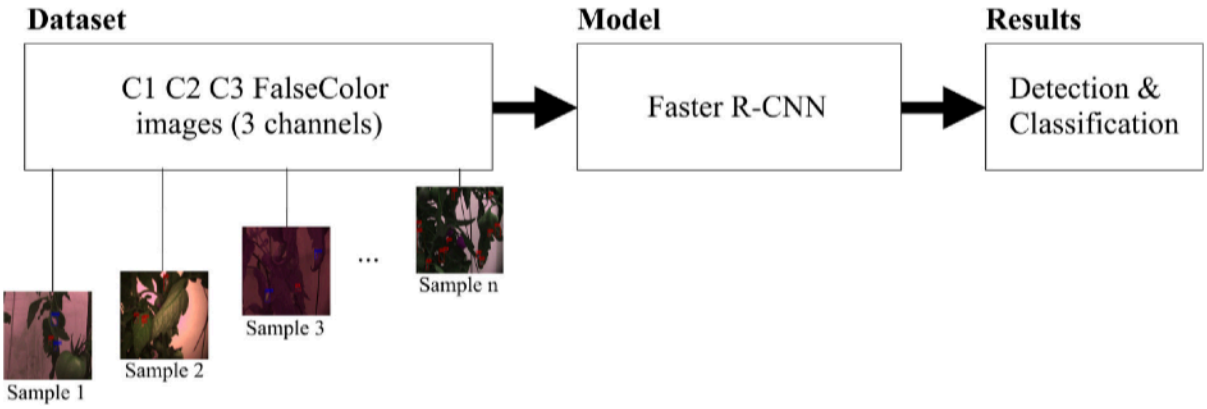


Fig. 6. Pipeline of the reconstructed 3-channel method, where images combining different channels (i.e. spectral bands) are used.

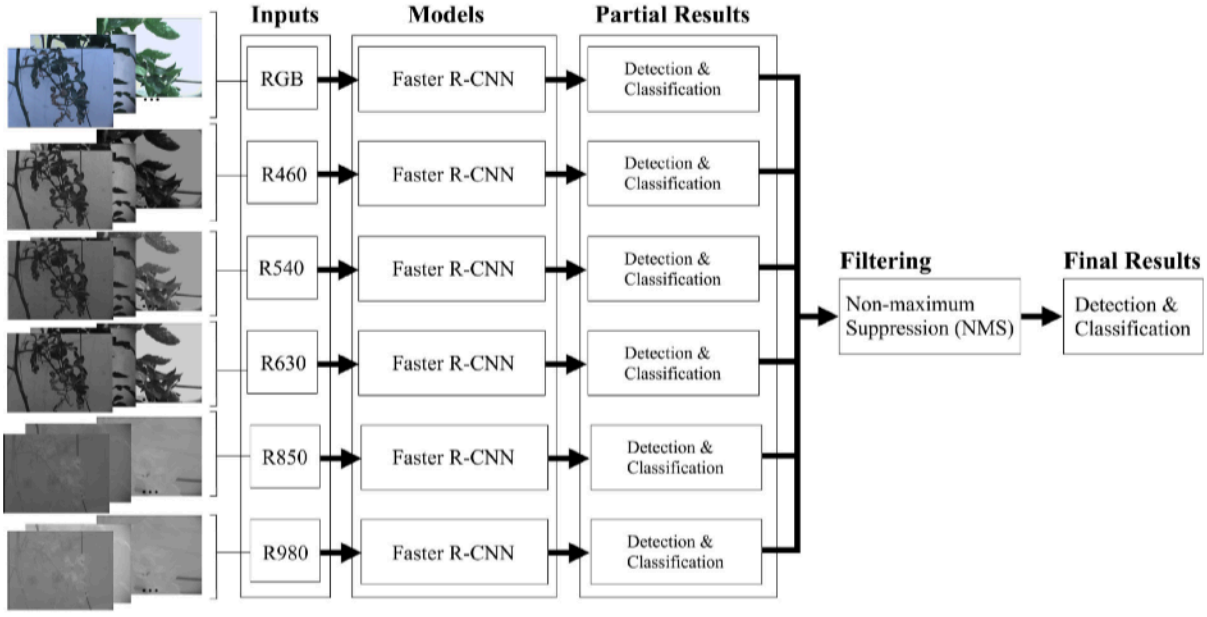


Fig. 7. The proposed MME approach is an adaptive fusion pipeline that integrates multispectral information from different spectral bands efficiently.

the final prediction through the Non-Maximum Suppression (NMS) technique (Felzenszwalb et al., 2008), as illustrated in Fig. 7.

NMS is a post-processing technique widely used in object detection to eliminate duplicate detections and select the most relevant bounding boxes corresponding to the detected objects. This approach allows one or more spectral bands to be exploited, individually or in combination, to improve prediction accuracy in plant disease detection. Each model learns the characteristics of the data independently, avoiding the need to reduce the number of channels to three or to use a single model.

In the proposed methodology, the input data may include various types of images from multi-modal acquisition, such as RGB, multi-spectral (VIS-NIR), thermal, as well as reconstructed images, such as false-colour, vegetation indices, etc. The employed architectures are flexible, ranging from individual models like Faster R-CNN, YOLO, etc., to combinations of these models within the same pipeline. Techniques such as Weighted Boxes Fusion (WBF) can be integrated alongside NMS to refine the most relevant predictions. These strategies are geared towards maximising the overall predictive performance.

In summary, using independent models for detection and leveraging multidimensional data enables more accurate final predictions. To our knowledge, this is the first approach applied to multispectral data for plant disease detection, validated in a real-world scenario. Furthermore, this methodology is inherently scalable and can be extended to other application domains.

2.4. Performance evaluation

For each method, we compute the standard Precision, Recall, F1-score, Average Precision (AP), and mean Average Precision (mAP) metrics to evaluate the efficiency of object detection. The class-wise metrics are based on the ratio of negative to positive; i.e, true positive (TP), and true negative (TN) instances that have been correctly classified and the number of positive and negative incorrectly classified, false positive (FP), and false negative cases (FN)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{and} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{2}$$

The F1-Score represents the harmonic mean of precision and recall into a single metric to understand model performance better

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

Intersection over Union (IoU) is the ratio of the intersection of the two boxes' areas to their combined areas. The ground truth bounding box and the anticipated bounding box both encompass the area of union, which is the denominator. We calculate the overlap between the ground-truth bounding box (A) and the predicted bounding box (B) in the numerator

$$\text{IoU} = \frac{\text{A} \cap \text{B}}{\text{A} \cup \text{B}}. \tag{4}$$

Table 2
Results of the Faster R-CNN model based on ResNet50 backbone modified through Conv2D to accept input images with more than 3 channels. The mAP appears in bold; the scores for each pathology class are also provided. Scores for mAP, AP, Precision, Recall and F1-score refer to %.

Dataset	Batch size	mAP	AP		Precision		Recall		F1-score	
			Tuta	Oidium	Tuta	Oidium	Tuta	Oidium	Tuta	Oidium
Multi-channels method										
Full Stack	5	**46.99**	24.68	69.32	41.63	80.00	36.10	70.59	38.67	75.00
Stack S1	5	**49.97**	24.93	75.03	45.50	72.22	37.76	76.47	**41.27**	74.29
Stack S2	5	**50.95**	27.44	74.47	44.08	74.29	38.59	76.47	41.15	75.36
Stack S3	5	**50.08**	27.80	72.36	41.74	83.87	37.76	76.47	39.65	**80.00**
Stack S4	5	**48.19**	23.33	73.06	39.01	78.79	36.10	76.47	37.50	77.61
RGB method										
Original	1	**47.94**	25.28	70.61	41.12	71.43	36.51	73.53	38.68	72.46
Original	5	**46.81**	22.91	70.71	30.41	78.12	37.34	73.53	33.52	75.76
EBCC	1	**46.83**	23.33	70.34	39.90	75.76	31.95	75.53	35.48	74.63
EBCC	5	**45.03**	20.26	69.82	40.37	80.65	26.97	73.53	32.34	76.92
GreyWorld	1	**52.42**	23.98	**80.86**	37.55	75.68	36.93	**82.35**	37.24	78.87
GreyWorld	5	**46.83**	21.14	72.53	34.20	81.25	34.78	76.47	33.47	78.79
Spectral band method										
R460	5	**11.97**	8.46	15.48	28.67	35.29	17.01	17.65	21.35	23.53
R540	5	**35.00**	17.47	52.54	33.50	70.37	28.63	55.88	30.87	62.30
R630	5	**40.04**	20.24	59.84	34.38	77.78	31.95	61.76	33.12	68.85
R850	5	**4.26**	2.65	5.88	15.70	**100.0**	7.88	5.88	10.50	11.11
R980	5	**2.13**	1.32	2.94	8.33	14.29	6.22	2.94	7.13	4.88
Independent mixing method										
D1	5	**16.51**	6.78	26.24	45.89	82.19	9.27	29.41	15.42	43.32
D2	5	**25.01**	14.30	35.73	45.72	83.61	19.40	37.50	27.24	51.78
D3	5	**34.64**	12.91	56.38	36.96	74.07	21.16	58.82	26.91	65.57
D4	5	**40.05**	18.03	62.09	40.78	80.00	26.14	64.71	31.86	71.54
D5	5	**39.14**	19.58	58.72	**50.00**	77.36	25.93	60.29	34.15	67.77
Reconstructed 3-channels method										
FC1	5	**41.50**	20.55	62.45	35.21	62.16	31.12	67.65	33.04	64.79
FC2	5	**44.04**	21.49	66.60	33.61	79.31	33.20	67.65	33.40	73.02
FC3	5	**39.29**	20.68	57.91	33.05	67.74	32.37	61.76	32.70	64.62
FC4	5	**47.90**	25.29	70.53	40.19	71.43	35.68	73.53	37.80	72.46
FC5	5	**47.22**	**28.60**	65.84	34.47	85.19	41.91	67.65	37.83	75.41
FC6	5	**48.27**	24.16	72.39	30.69	78.79	38.59	76.47	34.19	77.61
NDVI	5	**39.82**	20.06	59.59	36.51	84.00	28.63	61.76	32.09	71.19
Multi-Model Ensemble method										
M1	5	**31.32**	16.63	46.02	11.89	30.11	**43.57**	**82.35**	18.68	44.09
M2	5	**32.68**	17.91	47.47	16.80	35.90	43.15	**82.35**	24.19	50.00
M3	5	**38.94**	17.22	60.67	25.07	50.91	37.34	**82.35**	30.00	62.92
M4	5	**37.68**	20.65	54.72	23.84	49.02	40.66	73.53	30.06	58.82

Average Precision (AP) is computed as the mean of the precision values (Precision(r_i)) at discrete recall points (r_i), where N represents the total number of recall points considered in the calculation

$$\mathrm{AP} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{Precision}(r_i). \tag{5}$$

The mean Average Precision (mAP) is the mean of AP between the different categories or classes detected and summarises the performance of a detection model

$$mAP = \frac{1}{N}\sum_{k=1}^{N} AP_k. \tag{6}$$

3. Results

All results are reported in Table 2. Hereby, the different experiments are discussed.

3.1. Multi-channels method

In Table 2, the results are reported by considering the entire multispectral stack as a single 8-channel image (Full Stack), as well as by following the stacks of RGB and subgroups of spectral bands (S1, S2, S3, S4) described previously. The metrics show that the *Tuta* class has significantly lower results than the *Oidium* class. This discrepancy is due to numerous false positives in the background, which share similar visual characteristics to *Tuta*, probably due to the non-optimal image acquisition conditions.

Analysing the mAP using the entire image stack or many spectral bands (S1 and S2) leads to inferior performance. Similarly, excessively reducing the number of available bands (S3 and S4) does not increase the model's accuracy. Instead, the most balanced compromise in terms of performance is obtained by combining RGB images with the most informative multispectral bands, also demonstrated by the subsequent analysis, as evidenced in the case of S2.

Fig. 8 shows some examples of disease detection. As can be observed, most of the diseases were correctly identified with a high confidence score. On the other side, there are some cases located in the background, therefore slightly out of focus or in a darker area, which are not identified. In this context, the main objective is to improve the scalability of the method, maximising the capacity to identify as many diseases as possible, especially in the early stages. In addition, the first example shows a failure to identify *Oidium* in the early stages of disease development, highlighting the difficulties for detection using this method. Consequently, further analyses are conducted to refine the predictions and explore additional strategies to improve the accuracy and robustness of detection.

Fig. 8. Results output of the multi-channels method using full stack images (8-channels, a batch size of 5 and 30 epochs). The colour red is used to mark the Tuta and blue the Oidium for recognition.

3.2. RGB method

The results in Table 2 were obtained by considering the original RGB images and, subsequently, those optimised using two image enhancement approaches: EBCC and GreyWorld. The complete analysis is reported in Table 4 in Appendix.

The mAP results show that good learning results are obtained even if only RGB images are considered. However, it should be noted that several false positives are observed in the original images due to the non-uniform yellowish-grey background usually associated with *Tuta*. These false positives were reduced in the datasets with the enhanced images, in which the yellow tones in the background were smoothed out, minimising similarities with leaf disease characteristics, as shown in Fig. 9. Although the most representative class of *Tuta* continues to perform significantly low for all evaluation metrics.

In addition, we observed that some predictions classified as false positives may represent true positives that were inaccurately annotated. This discrepancy likely arises partly from the inherent limitations of human visual inspection, where subtle features may be overlooked due to limited attention to detail and partly from increased error rates associated with observer tiredness.

3.3. Spectral band method (single-channel)

Table 2 reports the results related to the analysis obtained from the individual spectral bands (R460, R540, R630, R850, and R980) using Faster R-CNN with the ResNet50 backbone. The full analysis is reported in Table 5 in Appendix.

We observed that the R540 and R630 spectral bands produced the best performance in terms of detection, while the R460 band showed poor results, and the NIR bands, R850 and R980 bands showed significantly lower performance, as illustrated in Figs. 10. This behaviour can be explained by analysing the nature of the symptoms associated with the two tomato diseases (Tuta absoluta and Leveillula taurica), which manifest themselves as colour changes in shades of yellow, orange and brown. In particular, the R540 band is effective because, in the presence of the disease, there is a decrease in reflectance due to the degradation of the chlorophyll, while healthy leaves tend to reflect this wavelength more, returning a visual appearance perceived as green. Conversely, an increase in reflectance in the R630 band is associated with the appearance of yellow-orange pigmentation on infected leaves, making it particularly informative for early detection. Regarding the NIR bands, the results obtained with R850 and R980 were lower than those obtained with the visible bands. This can be attributed to the fact that the pathologies analysed produce visible symptoms mainly in the form of superficial changes in leaf aspect, which are mainly manifested in the visible spectrum. Although NIR imaging is generally effective for analysing physiological parameters such as water content, internal tissue structure or biomass index, it is less sensitive to the surface pigment changes that characterise the early stages of leaf diseases. As a result, NIR bands provide limited information to object detection-based detection models, which rely heavily on visible spectral contrast to identify and locate affected areas accurately. This is reflected in an overall decrease in the precision and recall metrics associated with these bands.

Segmentation of *Tuta* shows much worse results than *Oidium*, indicating that the acquired bands are unsuitable for accurately characterising the manifestation of the *Tuta*. This aspect underlines the need to identify spectral signatures and acquire imaging in the bands that better characterise such phenomena.

3.4. Independent mixing method

Table 2 reports the results of the combined analysis of RGB and multispectral images using all bands or only those with higher learning scores obtained from the previous analyses as independent samples. The dataset consisting of RGB and spectral images, considering all data, is called D1 or, in subgroups, as D2, D3, D4 and D5, described in Section 2.3.

As expected, considering RGB images with all spectral bands as separate examples for models leads to lower learning scores than those with a combination of RGB and bands with higher scores from previous analyses (R540 and R630). Moreover, this method has not improved performance compared to previous methods. So, using a single model with data with different spectral information leads to more confusion than helping to extract useful features. This aspect motivated us to explore further methods.

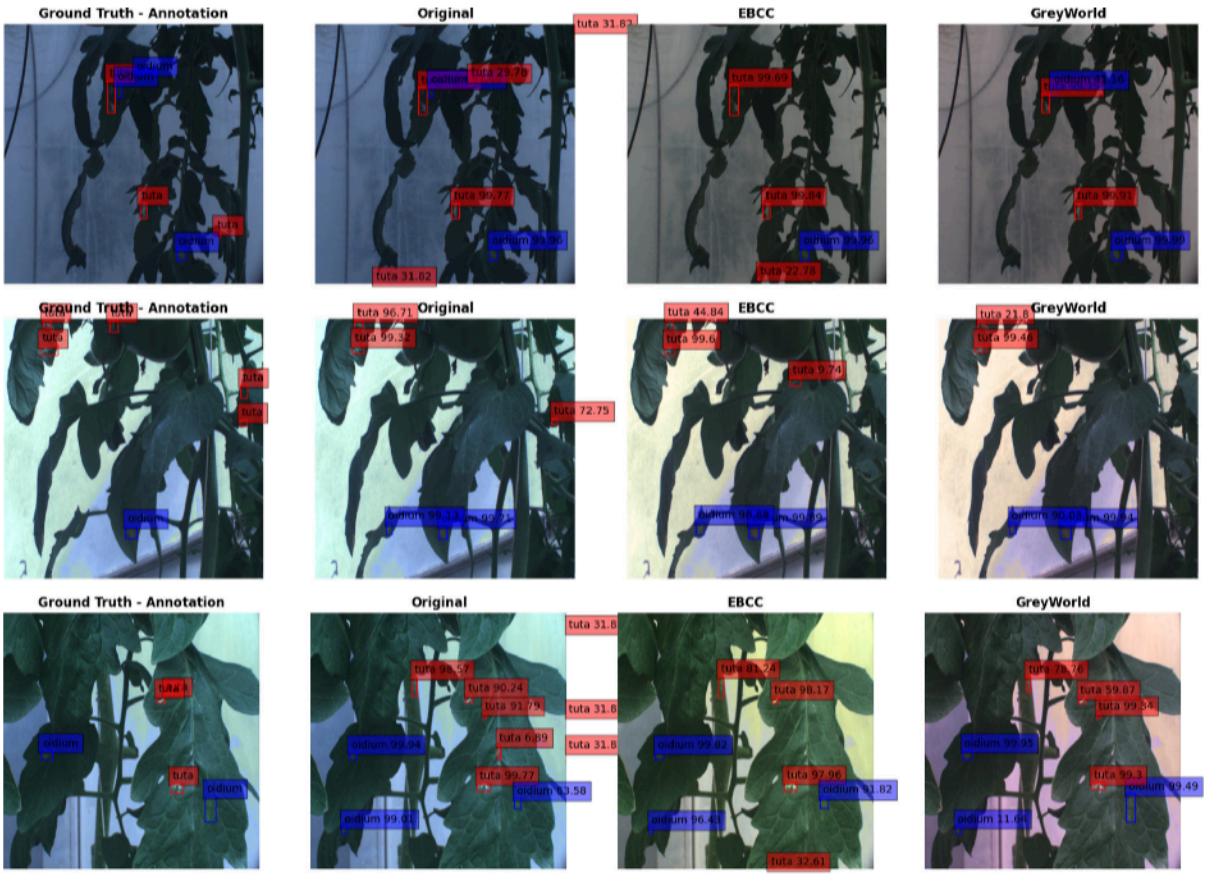


Fig. 9. Results output of the RGB method with models trained on original and enhanced RGB images with a batch size of 1 and 30 epochs.

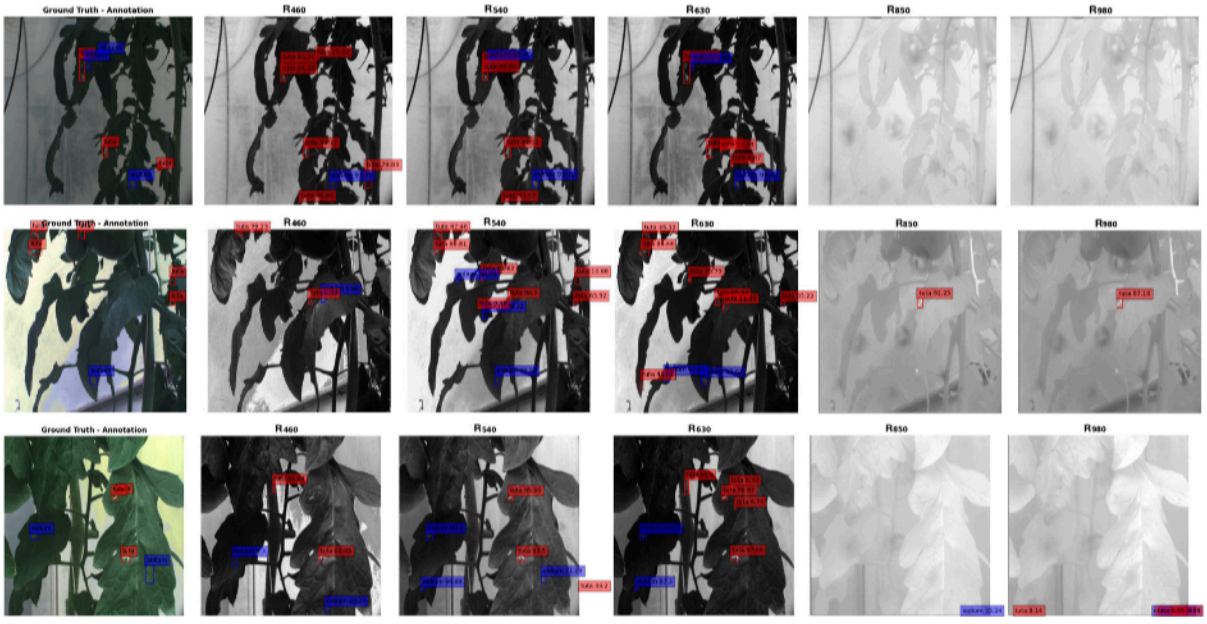


Fig. 10. Results output of the spectral band (single-channel) method, where models was trained with a batch size of 5 and 30 epochs.

3.5. Reconstructed 3-channel method

Table 2 shows the results of analysing combinations of spectral images with RGB images by replacing one or more channels of the RGB images with spectral bands, resulting in false-colour images, called FC1, FC2, FC3, FC4 and FC5 or as a spectral bands combination called FC6, described in Section 2.3. In addition, we also use the NDVI as input data.

The mAP metrics show that combining the spectral bands leads to superior performance compared to analysing each band independently. Although the improvement is not drastically more significant than the initial methods, which were already performing well, this method

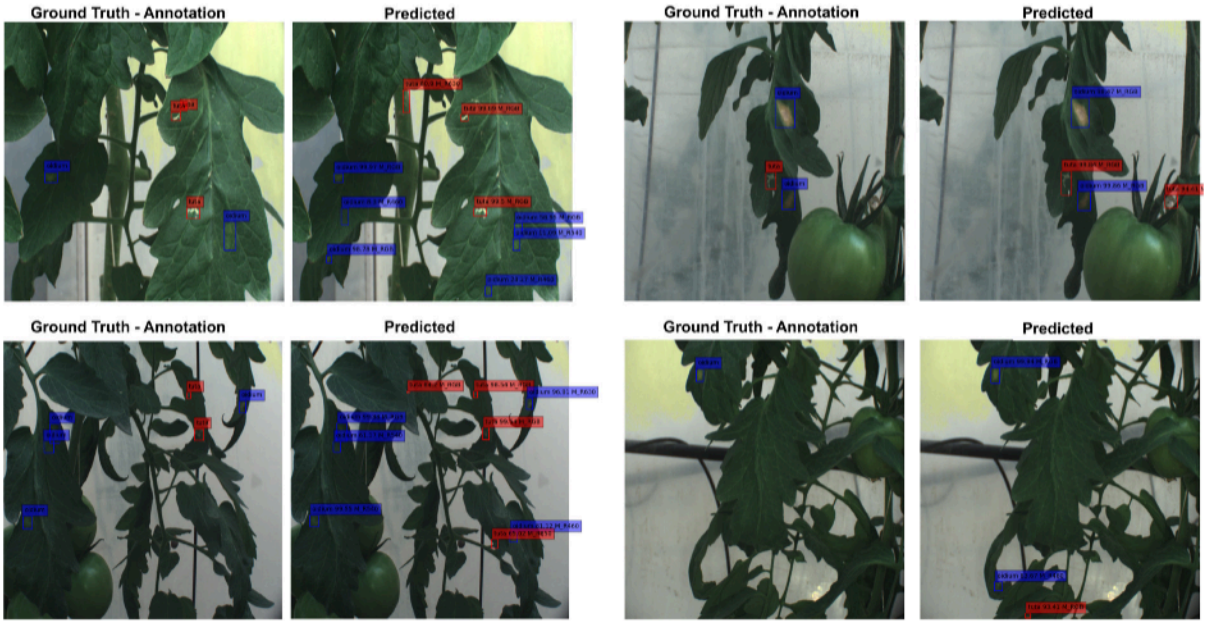


Fig. 11. Results output of our MME pipeline with the complete multispectral dataset M1.

shows excellent potential as it reduces false positives associated with the background. This is probably due to the increased discriminative ability to integrate spectral information, which allows relevant features to be more accurately distinguished from background signals. These results underline the importance of spectral data fusion to increase the robustness and reliability of detection models, especially in complex environments. This led to the design of the novel MME pipeline with a partial results fusion block to merge the predictions and select the most valuable predictions via NMS for a more precise and accurate final result.

3.6. Multi-model ensemble method

Table 2 summarises the results obtained from the different configurations. The mean mAP of the MME is slightly lower than that of some individual models, which is expected due to the aggregation of multiple predictions and the presence of both false positives and false negatives. However, the ensemble produces more stable and consistent outputs, reducing background-related errors and highlighting unlabelled regions that likely correspond to true disease instances at an early stage. This suggests a higher generalisation capability and robustness to imperfect or inconsistent annotations, aspects that conventional metrics do not fully capture.

Fig. 11 illustrates examples of the MME outputs, where most annotated instances were correctly detected and classified. Some of the predictions initially labelled as false positives may actually correspond to early symptoms of disease, visible in leaf areas not included in the original annotations. These findings confirm the effectiveness and reliability of the proposed approach, although further validation on larger and more diverse datasets is required, particularly for early-stage disease detection, where visual identification is inherently challenging.

4. Discussion

The proposed MME pipeline allows to handling different multispectral configurations and it has been demonstrated in an agriculture case study. We analysed and compared different multispectral data processing and integration approaches to identify optimal channel selection and method strategies that maximise the learning effectiveness. The training (and so the inference) of deep CNN depends on the structure of the input channels, which significantly affects the initial feature extraction and the learning process. An RGB image or a three-channel false-colour image is processed differently from individual channels considered separately, as the model learns the relevant features using convolutional layers by applying in one step a set of filters to all three channels to produce a single feature map. While, in the case of multiple monochromatic image inputs the CNN produces multiple features maps resulting more informative.

In the work by Johnson et al. (2021) it is demonstrated that the heterogeneous inputs affect how the model learn robust spatio-temporal representations. For example, converting images to alternative colour spaces facilitates the extraction of discriminative features that are less sensitive to environmental and lighting variations. These results highlight the importance of dataset composition to learn more relevant features, especially in complex and noisy scenarios. Anyway, not all multispectral bands are equally informative; treating the entire multispectral stack as a single input may introduce redundancy or noise, compromising the learning phase.

The results highlight several critical aspects and points for improvement in early disease detection using multispectral images and DL. In particular, the metrics indicate that the class related to *Tuta* performs significantly poorer than that of *Oidium* due to the presence of numerous false positives in the background, which share similar visual characteristics to the disease, due to non-optimal image acquisition conditions. In addition, several samples exhibit out-of-focus regions, resulting in difficulties for both detection and annotation.

The mAP metrics revealed that the multi-channel method, when using either the entire multispectral stack or too many bands (S1 and S2 configurations), decreases the performance; on the other hand, an excessive reduction of the number of bands (S3 and S4), does not allow spectral information to be fully exploited. The most significant trade-off in performance is obtained by combining RGB images with the most informative multispectral bands, as evidenced by the S2 configuration. This aspect highlights the importance of carefully selecting spectral bands to acquire data with relevant information, especially in the case of early disease detection in plants (Scutelnic et al., 2024b). This spectral characterisation will make it possible to optimise the optical chain and thus the imaging, reducing costs, the data to be acquired and the computing power required without decreasing performance.

Table 3
Comparative summary of recent studies using multispectral data for plant disease detection. mAP and Accuracy values are expressed in %. "–" indicates not reported.

Study	Data type	Approach	mAP/Accuracy (%)	Crop/Disease	Environment	Contribution
Georgantopoulos et al. (2023)	Multispectral + RGB	Faster R-CNN with feature fusion	mAP 18.5–20.2/F1 > 90	Tomato (T. absoluta, L. taurica)	Greenhouse	Public dataset and multispectral detection baseline
Giakoumoglou et al. (2023)	RGB	Object detection with ensemble (WBF, NMS)	mAP 58–70	Tomato (T. absoluta)	Green-house/Field	Improved mAP using ensemble detection
De Silva and Brown (2023a)	RGB + NIR	CNN (MobileNet, ResNet-50V2) and ViT	Accuracy up to 98	Mixed crops	Natural conditions	Spectral bands boost accuracy over RGB
Kerkech et al. (2020)	Multispectral + Depth Map	VddNet segmentation network	–	Vine	Field	Deep segmentation combining spectral and depth
This study	Multispectral + RGB	MME object detection pipeline	**mAP 38.94**	Tomato (T. absoluta, L. taurica)	Applied Greenhouse	Adaptive fusion and spatial detection from spectral inputs

The comparison between the original RGB and images obtained with enhancement techniques (EBCC and GreyWorld) showed a reduction in false positives. However, the *Tuta* class performed lower, underlining the importance of further optimisation in the acquisition and pre-processing phases. It is important to note that images provide physical information, and if some of them is lost during acquisition, even the most advanced algorithms cannot accurately reconstruct real phenomena. This is one of the main limitations to consider, especially in complex operational contexts such as imaging in an open field or in low light conditions that increase the difficulty of acquisition and analysis.

The individual analysis of the spectral bands confirmed that the R540 and R630 bands provide the best results, while the R460 band shows bad performance, and the R850 and R980 bands perform poorly. These results are consistent with the physical properties associated with the disease: reduced reflectance of chlorophyll in the R540 band and increased reflectance in the R630 band, which manifests itself in the yellow-orange tones and brown hues observed in these two diseases.

The mixing method, combining of RGB and all spectral images (D1) shows low learning ability: different spectral inputs confuse the model rather than help it. By progressively reducing the number of bands (D2, D3, D4 and D5) to those significant for diagnosis, an improvement in learning is noted, but with a mAP no higher than that of the previously analysed methods.

Combining the spectral bands as a reconstructed 3-channel method improves performance compared to analysing each band separately, reducing false positives due to background. Although enhancing the initial methods is not radical, integrating spectral information increases the discriminative capacity, strengthening the robustness and reliability of the models, especially in complex contexts.

Finally, the proposed method, based on the MME, highlights the potential of a hybrid approach to exploit the full dimensionality of the data: independent models generate predictions, and NMS removes duplicates to achieve a more accurate final prediction with a higher confidence score. Unlike traditional fusion strategies that rely on feature-level concatenation or fixed-band selection, the MME architecture preserves the full multispectral diversity by operating at the decision level. This design enables the model to integrate data with opposite physical meanings, such as reflectance and absorbance, or even emissivity information from thermal sensors. Consequently, it can process heterogeneous datasets without requiring spectral homogenisation or dimensionality reduction, a capability not reported in prior literature. The results obtained support the use of ensemble-based methods and multispectral data as promising tools for plant disease monitoring. However, the aggregation of results resulted in a slight reduction in mAP, partly due to the quality of the data and partly due to false negatives, which we suspect to be true positives not annotated due to the limitations of the human eye, attention to details and tiredness. This suggests the need for further studies and the use of multispectral datasets with temporal data to evaluate the proposed method further.

Recent studies (reported in Table 3) have explored the use of multispectral data and DL techniques for the early detection of diseases in tomato plants and other crops. Georgantopoulos et al. (2023) proposed a multispectral dataset comprising 5 multispectral bands plus RGB, aimed at detecting Tuta absoluta and Leveillula taurica in greenhouse environments. Their approach uses a Faster R-CNN with band fusion, achieving a F1-score > 0.9 and a mAP between 18.5% and 20.2%. Giakoumoglou et al. (2023) applied ensemble techniques to object detection models on real RGB datasets acquired both in greenhouses and in open fields, achieving an improvement in mAP from 0.58 to 0.70 using Weighted Boxes Fusion. A relevant study is that of De Silva and Brown (2023a), which addressed the classification of multispectral images of diseased leaves in natural environments by combining CNN and vision transformers (ViT) to improve accuracy. Their work focused on comparative analysis between RGB, NIR and multispectral data, demonstrating that the integration of different spectral bands, particularly those covering both the visible and NIR spectrum, produces a significant improvement in classification accuracy compared to the use of RGB channels alone. This result suggests that combining adjacent spectral regions can better highlight the symptoms associated with different diseases, even in the early stages. Kerkech et al. (2020) developed VddNet, an architecture that uses multispectral images and depth maps to detect vine diseases, offering spatial segmentation and robustness in real environments. Although these studies focus on classification rather than object detection, they reinforce our choice to design a pipeline that makes targeted use of spectral selection and adaptive fusion, avoiding arbitrary reductions in dimensionality and fully exploiting the informational potential of multispectral images.

Based on these results, our work adopts a strategy that is not limited to classification but fully exploits the multidimensionality of spectral information through object detection techniques. We have designed a pipeline based on an MME approach for object detection, which integrates different spectral fusion and selection strategies, achieving better performance and more accurate localisation of infections compared to previous methods. However, the generalisability of the results will need to be verified with larger datasets, different plant species and variable acquisition conditions, aspects that represent the natural continuation of this work.

4.1. Limitations and future work

Despite the promising results obtained through the proposed approach based on a MME pipeline, it is worth highlighting some limitations. First, the open-source dataset used for training and validation,

Table 4
Results of the Faster R-CNN detector evaluation with the ResNet50 backbone with the dataset consisting only of original and optimised RGB images. The average accuracy appears in bold; two pathology classes organise the other scores. Scores for mAP, AP, Precision, Recall and F1-score refer to %.

Dataset	Batch size	Epochs	mAP	AP		Precision		Recall		F1-score	
				Tuta	Oidium	Tuta	Oidium	Tuta	Oidium	Tuta	Oidium
Original	1	30	**47.94**	25.28	70.61	41.12	71.43	36.51	73.53	38.68	72.46
Original	1	50	**47.76**	24.37	71.15	44.51	78.79	31.95	76.47	37.20	77.61
Original	1	100	**36.35**	16.10	56.60	46.34	76.92	23.65	58.82	31.32	66.67
Original	5	30	**46.81**	22.91	70.71	30.41	78.12	37.34	73.53	33.52	75.76
Original	5	50	**49.85**	24.27	75.44	39.71	71.79	34.44	82.35	36.89	76.71
Original	5	100	**49.79**	20.49	79.10	41.71	82.35	30.29	82.35	35.10	82.35
EBCC	1	30	**46.83**	23.33	70.34	39.90	75.76	31.95	75.53	35.48	74.63
EBCC	1	50	**36.82**	20.02	53.64	40.70	70.37	29.05	55.88	33.90	62.30
EBCC	1	100	**43.32**	20.89	65.76	52.42	82.14	30.29	67.65	38.42	74.19
EBCC	5	30	**45.03**	20.26	69.82	40.37	80.65	26.97	73.53	32.34	76.92
EBCC	5	50	**50.14**	23.10	77.19	41.43	61.70	36.10	85.29	38.58	71.60
EBCC	5	100	**50.42**	23.36	77.49	34.60	66.67	34.02	82.35	34.31	73.68
EBCC	10	30	**45.40**	19.39	71.43	38.04	70.27	29.05	76.47	32.94	73.24
EBCC	10	50	**46.18**	21.67	70.70	37.62	70.27	31.54	76.47	34.31	73.24
EBCC	10	100	**45.27**	24.84	65.70	40.31	77.42	32.78	70.59	36.16	73.85
EBCC	15	30	**44.06**	21.04	67.09	34.77	70.59	36.93	70.59	35.81	70.59
EBCC	15	50	**47.78**	25.22	70.35	33.72	70.27	36.51	76.47	35.06	73.24
EBCC	15	100	**48.90**	26.44	71.36	34.55	78.79	39.42	76.47	36.82	77.61
EBCC	20	30	**43.44**	19.95	66.93	28.57	77.42	34.01	70.59	31.06	73.85
EBCC	20	50	**46.46**	23.40	69.52	38.22	71.43	34.02	73.53	36.04	72.46
EBCC	20	100	**45.70**	24.09	67.31	35.37	72.73	36.10	70.59	35.73	71.64
GreyWorld	1	30	**52.42**	23.98	80.86	37.55	75.68	36.93	82.35	37.24	78.87
GreyWorld	1	50	**46.07**	23.02	69.13	48.45	85.71	32.37	70.59	38.81	77.42
GreyWorld	1	100	**37.81**	13.98	61.64	49.51	78.57	21.16	64.71	29.65	70.97
GreyWorld	5	30	**46.83**	21.14	72.53	34.20	81.25	34.78	76.47	33.47	78.79
GreyWorld	5	50	**46.86**	23.20	70.54	41.21	78.12	31.12	73.53	35.46	75.76
GreyWorld	5	100	**49.66**	25.08	74.24	40.27	77.14	36.93	79.41	38.53	78.26

although carefully selected and annotated, is limited in size and variability. This could affect the models' ability to generalise to other plant species, different diseases or more complex environmental conditions. In particular, the differences in performance observed between the two diseases analysed, Tuta absoluta and Leveillula taurica, highlight the importance of disease-specific spectral signatures and the sensitivity of the detection process to background noise and image quality. A further limitation is the absence of temporal data. The progression of plant diseases is a dynamic phenomenon, and the initial symptoms may not be visible or distinguishable either visually or spectrally in a single acquisition. The integration of multispectral images acquired at different times would allow for more accurate modelling of disease evolution, thus improving the predictive power of the models. From a methodological point of view, although the ensemble approach has shown an improvement in detection capability by integrating the strengths of independent models, it also introduces an increase in computational load and requires careful calibration of fusion mechanisms (such as the NMS threshold), potentially limiting its applicability in real time on embedded or low-power devices.

Further research should focus on expanding and diversifying datasets to validate the proposed approach by testing it across different applications. The MME method is also applicable to other fields, as it can leverage various types of data with different physical meanings, such as UV, VIS, NIR, Short-Wave Infrared (SWIR), Long-Wave Infrared (LWIR or thermal), etc., to achieve more accurate results and higher confidence scores compared to using only visible bands, as in the case of RGB images.

5. Conclusions

In this study, we conducted a comparative analysis of state-of-the-art deep learning strategies for multispectral classification and detection of plant diseases, while also proposing a scalable approach based on a Multi-Model Ensemble architecture founded on a decision-level fusion strategy.

The results demonstrate that spectral information provides a substantial contribution to performance improvement compared with models trained exclusively on RGB data. Models trained on individual multispectral bands, particularly at wavelengths R540 and R630, achieved superior performance relative to the other bands, with mAP values of approximately 35.0% and 40.0%, respectively. The RGB baseline configuration, enhanced through GreyWorld normalisation, achieved an mAP of 47.9%. The integration of multispectral information through false-colour compositions (FC4, FC5, FC6) further improved performance, reaching an mAP of 48.3% and an F1-score of nearly 77%, thereby confirming the effectiveness of spectral fusion strategies.

The comparative analysis also revealed that not all spectral bands contribute equally to the classification process. The visible bands (R540 and R630) proved to be the most discriminative, as they are sensitive to physiological and pigment-related variations associated with disease symptoms. Conversely, the near-infrared bands (R850 and R980) exhibited limited relevance during the early stages of infection, with mAP values below 10%. The independent combination of RGB channels with the most informative multispectral bands (R540 and R630) showed that selective integration, based on the most significant spectral components, can enhance model generalisation while simultaneously reducing data redundancy. The D4 and D5 configurations derived from this strategy achieved an mAP of around 40% and an F1-score of 71%, demonstrating that even a limited yet well-chosen spectral subset can deliver competitive results.

The MME approach, based on the decision-level combination of outputs from specialised models using Non-Maximum Suppression, ultimately provided more robust and generalisable performance. The most effective configuration (M3: RGB + R540) achieved an mAP of 38.9% and an F1-score of 62.9%. Although the mean mAP was slightly lower than that of some single models, the ensemble produced more stable and consistent predictions, reducing background-related false positives and correctly identifying unlabelled instances, presumably true positives corresponding to early stages of disease development.

Beyond numerical metrics, and from an application perspective, the main advantage of the proposed MME pipeline lies in its high

Table 5
Results of the evaluation of the Faster R-CNN detector with the ResNet50 backbone. The dataset consists of monochrome images, taking each spectral band independently (460 nm, 540 nm, 630 nm, 850 nm, and 980 nm). Scores for mAP, AP, Precision, Recall, and F1-score refer to %.

Dataset	Batch size	Epochs	mAP	AP		Precision		Recall		F1-score	
				Tuta	Oidium	Tuta	Oidium	Tuta	Oidium	Tuta	Oidium
R460	1	30	**6.83**	8.76	4.90	34.78	50.00	13.28	5.88	19.22	10.53
R460	5	30	**11.97**	8.46	15.48	28.67	35.29	17.01	17.65	21.35	23.53
R540	1	30	**34.73**	13.84	55.63	38.00	50.00	23.65	6176	29.16	55.26
R540	5	30	**35.00**	17.47	52.54	33.50	70.37	28.63	55.88	30.87	62.30
R630	1	30	**35.39**	15.20	55.59	34.59	82.61	22.82	55.88	27.50	66.67
R630	5	30	**40.04**	20.24	59.84	34.38	77.78	31.95	61.76	33.12	68.85
R850	1	30	**3.89**	2.88	4.90	14.98	50.00	12.86	5.88	13.84	10.53
R850	5	30	**4.26**	2.65	5.88	15.70	100.00	7.88	5.88	10.50	11.11
R980	1	30	**2.23**	1.52	2.94	11.67	33.33	5.81	2.94	7.76	5.41
R980	5	30	**2.13**	1.32	2.94	8.33	14.29	6.22	2.94	7.13	4.88

flexibility and scalability, with the ability to integrate spectrally and physically heterogeneous data sources (reflectance, absorbance, and emissivity) within a unified framework, without the need for dimensionality reduction. This makes it particularly suitable for multi-source and multimodal scenarios (VIS–NIR–SWIR–LWIR). Its robustness to spectral variability, illumination conditions, and annotation quality further supports its applicability in real-world precision agriculture contexts.

In summary, this work demonstrates that the MME approach provides a practical and generalisable solution for managing and integrating complex multispectral data, overcoming the limitations of traditional single-model architectures. Future research will focus on extending the validation to larger datasets and incorporating additional spectral domains, including thermal and temporal data, to enhance early-stage disease identification capabilities further.

CRediT authorship contribution statement

Dumitru Scutelnic: Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Claudia Daffara:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Funding acquisition, Formal analysis. **Riccardo Muradore:** Writing – review & editing, Visualization, Supervision, Resources, Formal analysis. **Martin Weinmann:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Data curation, Conceptualization. **Boris Jutzi:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, 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.

Uncited references

Brand et al. (2009)

Acknowledgements

The research was supported by the following grant: Decree of the Ministry of University and Research No. 1061 of August 10, 2021, for the allocation of FSE REACT-EU resources under the PON “Ricerca e Innovazione” 2014–2020.

The research of Riccardo Muradore has been carried out within the PNRR research activities of the consortium iNEST (Interconnected North-East Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) Missione 4 Componente 2, Investimento 1.5 D.D. 1058 23062022, ECS_00000043).

This manuscript reflects only the Author views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

The authors would like to acknowledge the financial support from the University of Verona for covering the publication costs.

Appendix

Table 4 lists all estimates made with the datasets of RGB images at different batch sizes and epochs.

Table 5 lists all estimates made with the datasets of monochromatic images acquired in different spectral bands at different batch sizes and epochs.

Data availability

The used dataset is available in literature: Georgantopoulos et al. (2023). The source code and weights of the trained models are available at the GitHub repository: https://github.com/DumitruScutelnic/MME.

References

Abade, André, Ferreira, Paulo Afonso, de Barros Vidal, Flavio, 2021. Plant diseases recognition on images using convolutional neural networks: A systematic review. Comput. Electron. Agric. 185, 106125.

Aegerter, B.J., Stoddard, C.S., Miyao, E.M., Le Strange, M., Turini, T.A., 2015. Impact of powdery mildew (Leveillula Taurica) on yield and fruit quality of processing Tomatoes in California. Acta Hortic. 1081, 153–158. http://dx.doi.org/10.17660/ActaHortic.2015.1081.17.

Almahasneh, M., Paiement, A., Xie, X., et al., 2022. MLMT-CNN for object detection and segmentation in multi-layer and multi-spectral images. Mach. Vis. Appl. 9, http://dx.doi.org/10.1007/s00138-021-01261-y.

Barros, T., Conde, P., Gonçalves, G., Premebida, C., Monteiro, M., Ferreira, C.S.S., Nunes, U.J., 2022. Multispectral Vineyard segmentation: A deep learning comparison study. Comput. Electron. Agric. 195, 106782.

Bhargava, A., Sachdeva, A., Sharma, K., Alsharif, M.H., Uthansakul, P., Uthansakul, M., 2024. Hyperspectral imaging and its applications: A review. Heliyon 10 (12), e33208. http://dx.doi.org/10.1016/j.heliyon.2024.e33208, Available at: https://www.sciencedirect.com/science/article/pii/S2405844024092399.

Bharman, Pallab, Saad, Sabbir Ahmad, Khan, Sajib, Jahan, Israt, Ray, Milon, Biswas, Milon, 2022. Deep learning in agriculture: A review. Asian J. Res. Comput. Sci. 2022, 28–47. http://dx.doi.org/10.9734/ajrcos/2022/v13i230311, https://journalajrcos.com/index.php/AJRCOS/article/view/260.

Brand, M., Messika, Y., Elad, Y., Rav David, D., Sztejnberg, A., 2009. Spray treatments combined with climate modification for the management of *Leveillula Taurica* in sweet pepper. Eur. J. Plant Pathol. 124 (2), 309–329. http://dx.doi.org/10.1007/s10658-008-9421-z.

Buchsbaum, G., 1980. A spatial processor model for object colour perception. J. Franklin Inst. 310 (1), 1–26.

Butte, S., Vakanski, A., Duellman, K., Wang, H., Mirkouei, A., 2021. Potato crop stress identification in aerial images using deep learning-based object detection. Agron. J. 113 (5), 3991–4002.

Daffara, C., Muradore, R., Piccinelli, N., Gaburro, N., de Rubeis, T., Ambrosini, D., 2020. A cost-effective system for aerial 3D thermography of buildings. J. Imaging 6 (8), 76. http://dx.doi.org/10.3390/jimaging6080076.

Daffara, C., Parisotto, S., Mariotti, P.I., Ambrosini, D., 2021. Dual mode imaging in mid infrared with thermal signal reconstruction for innovative diagnostics of the Monocromo by Leonardo da Vinci. Sci. Rep. 11 (1), 22482. http://dx.doi.org/10.1038/s41598-021-01837-8.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 886–893.

de Manincor, N., Marchioro, G., Fiorin, E., Raffaelli, M., Salvadori, O., Daffara, C., 2020. Integration of multispectral visible-infrared imaging and pointwise X-ray fluorescence data for the analysis of a large canvas painting by Carpaccio. Microchem. J. 153, 104469. http://dx.doi.org/10.1016/j.microc.2019.104469.

De Silva, Malithi, Brown, Dane, 2023a. Multispectral plant disease detection with vision transformer–convolutional neural network hybrid approaches. Sensors 23 (20), 8531. http://dx.doi.org/10.3390/s23208531.

De Silva, M., Brown, D., 2023b. Plant disease detection using multispectral imaging. Commun. Comput. Inf. Sci. 1528.

Desneux, N., Wajnberg, E., Wyckhuys, K.A.G., Burgio, G., Arpaia, S., Narváez-Vasquez, C.A., González-Cabrera, J., Catalán Ruescas, D., Tabone, E., Frandon, J., Pizzol, J., Poncet, C., Cabello, T., Urbaneja, A., 2010. Biological invasion of European Tomato crops by Tuta absoluta: ecology, geographic expansion and prospects for biological control. J. Pest Sci. 83 (3), 197–215. http://dx.doi.org/10.1007/s10340-010-0321-6.

Estrada, J.S., Demarco, R., Johnson, C.M., et al., 2025. A multi-spectral and hyperspectral image dataset for evaluating chemical traits and the water status of Avocado, Olive and grape through leaf dehydration under laboratory conditions. Sci. Rep. 15, 2973. http://dx.doi.org/10.1038/s41598-025-85714-8.

Felzenszwalb, P., McAllester, D., Ramanan, D.A., 2008. Discriminatively trained, multiscale, deformable part model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8.

Fernández, Claudio I., Leblon, Brigitte, Wang, Jinfei, Haddadi, Ata, Wang, Keri, 2021. Detecting infected cucumber plants with close-range multispectral imagery. Remote. Sens. 13 (15), 2948. http://dx.doi.org/10.3390/rs13152948.

Gao, Zongmei, Luo, Zhongwei, Zhang, Wen, Lv, Zhenzhen, Xu, Yanlei, 2020. Deep learning application in plant stress imaging: A review. AgriEngineering 2 (3), 430–446. http://dx.doi.org/10.3390/agriengineering2030029, https://www.mdpi.com/2624-7402/2/3/29.

Gegenfurtner, K.R., Weiss, D., Bloj, M., 2024. Color constancy in real-world settings. J. Vis. 24 (2), 1–22.

Georgantopoulos, P.S., Papadimitriou, D., Constantinopoulos, C., Manios, T., Daliakopoulos, I.N., Kosmopoulos, D., 2023. A multispectral dataset for the detection of Tuta absoluta and Leveillula Taurica in Tomato plants. Smart Agric. Technol. 4, 100146. http://dx.doi.org/10.1016/j.atech.2022.100146.

Ghaderi, S., Fathipour, Y., Asgari, S., Reddy, G.V.P., 2019. Economic injury level and crop loss assessment for tuta absoluta (Lepidoptera: Gelechiidae) on different Tomato cultivars. J. Appl. Entomol. 143 (5), 493–507. http://dx.doi.org/10.1111/jen.12628.

Giakoumoglou, N., Pechlivani, E.M., Frangakis, N., Tzovaras, D., 2023. Enhancing Tuta absoluta detection on Tomato plants: Ensemble techniques and deep learning. AI 4, 996–1009. http://dx.doi.org/10.3390/ai4040050.

Girshick, Ross, 2015. Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448, Available at: https://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html.

Girshick, Ross, Donahue, Jeff, Darrell, Trevor, Malik, Jitendra, 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 580–587. http://dx.doi.org/10.48550/arXiv.1311.2524.

Guzman-Plazola, R.A., Davis, R.M., Marois, J.J., 2003. Effects of relative humidity and high temperature on spore germination and development of Tomato powdery mildew (Leveillula Taurica). Crop. Prot. 22 (10), 1157–1168. http://dx.doi.org/10.1016/S0261-2194(03)00157-1.

Halder, Moon, Datta, Ayon, Siam, Md Kamrul Hossain, Mahmud, Shakik, Sarkar, Md Saem, Rana, Md Masud, 2023. A systematic review on crop yield prediction using machine learning. In: The International Conference on Intelligent Systems & Networks. pp. 658–667. http://dx.doi.org/10.1007/978-981-99-4725-6_77, https://link.springer.com/chapter/10.1007/978-981-99-4725-6_77.

Hawkins, N.J., Bass, C., Dixon, A., Neve, P., 2019. The evolutionary origins of pesticide resistance. Biol. Rev. 94 (1), 135–155. http://dx.doi.org/10.1111/brv.12440, Available at: onlinelibrary.wiley.com.

Hughes, David, Salathé, Marcel, 2015. An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060.

Johnson, J., Sharma, G., Srinivasan, S., Masakapalli, S.K., Sharma, S., Sharma, J., Dua, V.K., 2021. Enhanced field-based detection of potato blight in complex backgrounds using deep learning. Plant Phenomics 2021, 9835724. http://dx.doi.org/10.34133/2021/9835724, PMID: 34104897; PMCID: PMC8147694.

Kerkech, M., Hafiane, A., Canals, R., 2020. VddNet: Vine disease detection network based on multispectral images and depth map. Remote. Sens. 12, 3305. http://dx.doi.org/10.3390/rs12203305.

Khan, Haider, Imran, Liu, Haiyan, Li, Wei, Cao, Aizhong, Wang, Xue, Liu, Hongyan, Cheng, Tao, Tian, Yongchao, Zhu, Yan, Cao, Weixing, et al., 2021. Early detection of powdery mildew disease and accurate quantification of its severity using hyperspectral images in wheat. Remote. Sens. 13 (18), 3612. http://dx.doi.org/10.3390/rs13183612.

Kulkarni, A.D., 2023. Multispectral image analysis using convolution neural networks. Int. J. Adv. Comput. Sci. Appl. (IJACSA) 14 (10), 1–5. http://dx.doi.org/10.14569/IJACSA.2023.0141002.

Kumar, Vinod, Singh, Ravi Shankar, Rambabu, Medara, Dua, Yaman, 2024. Deep learning for hyperspectral image classification: A survey. Comput. Sci. Rev. 53, 100658. http://dx.doi.org/10.1016/j.cosrev.2024.100658.

Leibe, B., Leonardis, A., Schiele, B., 2006. An implicit shape model for combined object categorization and segmentation. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (Eds.), Toward Category-Level Object Recognition. Springer Berlin Heidelberg, pp. 508–524.

Liu, Y., Meng, S., Wang, H., et al., 2024. Deep learning based object detection from multi-modal sensors: an overview. Multimedia Tools Appl. 83, 19841–19870. http://dx.doi.org/10.1007/s11042-023-16275-z.

Liu, J., Zhang, S., Wang, S., Metaxas, D.N., 2016. Multispectral deep neural networks for pedestrian detection. ArXiv, abs/1611.02644.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Lu, G., Fei, B., 2014. Medical hyperspectral imaging: A review. J. Biomed. Opt. 19 (1), 010901. http://dx.doi.org/10.1117/1.JBO.19.1.010901.

Lucas, J.A., Hawkins, N.J., Fraaije, B.A., 2015. Chapter two – the evolution of fungicide resistance. In: Sariaslani, S., Gadd, G.M. (Eds.), In: Advances in Applied Microbiology, Vol. 90, Academic Press, pp. 29–92. http://dx.doi.org/10.1016/bs.aambs.2014.09.001.

Manida, M., 2022. The Future of Food and Agriculture—Trends and Challenges. Agriculture & Food E-Newsletter, Available online: https://www.researchgate.net/publication/358402720_The_Future_of_Food_and_Agriculture_Trends_and_Challenges.

Musanase, Christine, Vodacek, Anthony, Hanyurwimfura, Damien, Uwitonze, Alfred, Kabandana, Innocent, 2023. Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. Agriculture 13 (11), 2141. http://dx.doi.org/10.3390/agriculture13112141, https://www.mdpi.com/2077-0472/13/11/2141.

Nowak, B., 2021. Precision agriculture: Where do we stand? A review of the adoption of precision agriculture technologies on field crops farms in developed countries. Agric. Res. 10 (4), 515–522. http://dx.doi.org/10.1007/s40003-021-00539-x.

Peng, Y., Dallas, M.M., Ascencio-Ibáñez, J.T., et al., 2022. Early detection of plant virus infection using multispectral imaging and spatial–spectral machine learning. Sci. Rep. 12, 3113. http://dx.doi.org/10.1038/s41598-022-06372-8.

Pesaresi, Simone, Mancini, Adriano, Quattrini, Giacomo, Casavecchia, Simona, 2024. Evaluation and selection of multi-spectral indices to classify vegetation using multivariate functional principal component analysis. Remote. Sens. 16 (7), 1224, https://www.mdpi.com/2072-4292/16/7/1224.

Pineda, Mónica, Barón, Matilde, Pérez-Bueno, María-Luisa, 2021. Thermal imaging for plant stress detection and phenotyping. Rem. Sens. 13 (1), 68. http://dx.doi.org/10.3390/rs13010068, Available at: https://www.mdpi.com/2072-4292/13/1/68.

Rayhana, R., Ma, Z., Liu, Z., Xiao, G., Ruan, Y., Sangha, J.S., 2023. A review on plant disease detection using hyperspectral imaging. IEEE Trans. AgriFood Electron. 1 (2), 108–134. http://dx.doi.org/10.1109/TAFE.2023.3329849.

Raza, S.E.A., Prince, G., Clarkson, J.P., Rajpoot, N.M., 2015. Automatic detection of diseased Tomato plants using thermal and stereo visible light images. PloS One 10 (4), e0123262.

Redmon, Joseph, Divvala, Santosh, Girshick, Ross, Farhadi, Ali, 2016. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 779–788, Available at: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html.

Ren, Shaoqing, He, Kaiming, Girshick, Ross, Sun, Jian, 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, Vol. 28, Available at: https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.

Ryckewaert, M., Héran, D., Trani, J.P., et al., 2023. Hyperspectral images of grapevine leaves including healthy leaves and leaves with biotic and abiotic symptoms. Sci. Data 10, 743. http://dx.doi.org/10.1038/s41597-023-02642-w.

Sa, Inkyu, Ge, Zongyuan, Dayoub, Feras, Upcroft, Ben, Perez, Tristan, McCool, Chris, 2016. DeepFruits: A fruit detection system using deep neural networks. Sensors 16 (8), 1222. http://dx.doi.org/10.3390/s16081222.

Saccuti, A., Graziosi, F., Lodi Rizzini, D., 2025. A dataset for vineyard disease detection via multispectral imaging. Data Brief. 61, 111712. http://dx.doi.org/10.1016/j.dib.2025.111712, PMID: 40534914; PMCID: PMC12173729.

Saleem, Muhammad Hammad, Potgieter, Johan, Arif, Khalid Mahmood, 2019. Plant disease detection and classification by deep learning. Plants 8 (11), 468, https://www.mdpi.com/2223-7747/8/11/468.

Salvi, Massimo, Acharya, U. Rajendra, Molinari, Filippo, Meiburger, Kristen M., 2021. The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. Comput. Biol. Med. 128, 104129. http://dx.doi.org/10.1016/j.compbiomed.2020.104129.

Santana, Oliverio J., Lorenzo-Navarro, Javier, Freire-Obregón, David, Hernández-Sosa, Daniel, Castrillón-Santana, Modesto, 2024. Applying deep learning image enhancement methods to improve person re-identification. Neurocomputing 598, 128011. http://dx.doi.org/10.1016/j.neucom.2024.128011.

Scutelnic, Dumitru, Muradore, Riccardo, Daffara, Claudia, 2024a. A multispectral camera in the VIS–NIR equipped with thermal imaging and environmental sensors for non-invasive analysis in precision agriculture. HardwareX 20, e00596. http://dx.doi.org/10.1016/j.ohx.2024.e00596.

Scutelnic, Dumitru, Muradore, Riccardo, Daffara, Claudia, 2024b. Multispectral imaging supervised by optical spectrometry for close acquisition in precision agriculture. In: Proceedings of the 2024 IEEE International Workshop on Metrology for Agriculture and Forestry. MetroAgriFor, pp. 598–602. http://dx.doi.org/10.1109/MetroAgriFor63043.2024.10948867.

Singh, Pal, Davinder, Jain, Naman, Jain, Pranjali, Kayal, Pratik, Kumawat, Sudhakar, Batra, Nipun, 2019. PlantDoc: A dataset for visual plant disease detection. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD.

Singh, V., Sharma, N., Singh, S., 2020. A review of imaging techniques for plant disease detection. Artif. Intell. Agric. 4, 229–242. http://dx.doi.org/10.1016/j.aiia.2020.10.002, Available at: https://www.sciencedirect.com/science/article/pii/S2589721720300295.

Tsouros, D.C., Bibi, S., Sarigiannidis, P.G., 2019. A review on UAV-based applications for precision agriculture. Information 10 (11), 349. http://dx.doi.org/10.3390/info10110349.

Tudi, M., Ruan, H.D., Wang, L., Lyu, J., Sadler, R., Connell, D., Chu, C., Phung, D.T., 2021. Agriculture development, pesticide application and its impact on the environment. Int. J. Environ. Res. Public Health 18 (3), 1112. http://dx.doi.org/10.3390/ijerph18031112, PMID: 33513796. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7908628/.

Vali, A., Comai, S., Matteucci, M., 2020. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. Remote. Sens. 12 (15), 2495. http://dx.doi.org/10.3390/rs12152495, https://www.mdpi.com/2072-4292/12/15/2495.

van de Weijer, J., Gevers, T., Gijsenij, A., 2007. Edge-based color constancy. IEEE Trans. Image Process. 16 (9), 2207–2214.

Vélez, S., Ariza-Sentís, M., Valente, J., 2023. Dataset on unmanned aerial vehicle multispectral images acquired over a vineyard affected by Botrytis Cinerea in northern Spain. Data Brief 46, 108876.

Wang, Q., Qi, F., Sun, M., Qu, J., Xue, J., 2019. Identification of tomato disease types and detection of infected areas based on deep convolutional neural networks and object detection techniques. Comput. Intell. Neurosci. 2019 (1), 9142753.

Wei, T., Chen, Z., Huang, Z., Yu, X., 2024a. Benchmarking in-the-wild multimodal disease recognition and a versatile baseline. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 1593–1601.

Wei, T., Chen, Z., Yu, X., Chapman, S., Melloy, P., Huang, Z., 2024b. Plantseg: A large-scale in-the-wild dataset for plant disease segmentation. arXiv preprint arXiv:2409.04038.

Weinmann, M., Weinmann, M., 2019. Fusion of hyperspectral, multispectral,color and 3D point cloud information for the semantic interpretation of urban environments. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. XLII-2/W13, 1899–1906.

Xu, H., Zhu, S., Ying, Y., Jiang, H., 2006. Early detection of plant disease using infrared thermal imaging. In: Optics for Natural Resources, Agriculture, and Foods, vol. 6381, SPIE, pp. 302–308. http://dx.doi.org/10.1117/12.685534, Available at: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/6381/638110/Early-detection-of-plant-disease-using-infrared-thermal-imaging/10.1117/12.685534.full.

Zhang, X., Vinatzer, B.A., Li, S., 2024. Hyperspectral imaging analysis for early detection of Tomato bacterial leaf spot disease. Sci. Rep. 14, 27666. http://dx.doi.org/10.1038/s41598-024-78650-6.

Zhang, N., Yang, G., Pan, Y., Yang, X., Chen, L., Zhao, C., 2020. A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades. Remote. Sens. 12 (19), 3188. http://dx.doi.org/10.3390/rs12193188, Available at: https://www.mdpi.com/2072-4292/12/19/3188.

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J., 2023. Object detection in 20 years: A survey. Proc. IEEE 111 (3), 257–276.