agriculture

MDPI

Article

Comparative Evaluation of AI-Based Multi-Spectral Imaging and PCR-Based Assays for Early Detection of *Botrytis cinerea* Infection on Pepper Plants

Dimitrios Kapetas [1], Eleni Kalogeropoulou [2], Panagiotis Christakakis [1], Christos Klaridopoulos [3] and Eleftheria Maria Pechlivani [1,*]

[1] Information Technologies Institute, Centre for Research and Technology Hellas, 57001 Thessaloniki, Greece; dimikape@iti.gr (D.K.); christakakis@iti.gr (P.C.)
[2] Laboratory of Mycology, Scientific Directorate of Phytopathology, Benaki Phytopathological Institute, 14561 Athens, Greece; e.kalogeropoulou@bpi.gr
[3] iKnowHow S.A., 15451 Athens, Greece; cklaridopoulos@iknowhow.com
* Correspondence: riapechl@iti.gr; Tel.: +30-231-125-7751

Abstract: Pepper production is a critical component of the global agricultural economy, with exports reaching a remarkable $6.9B in 2023. This underscores the crop's importance as a major economic driver of export revenue for producing nations. *Botrytis cinerea*, the causative agent of gray mold, significantly impacts crops like fruits and vegetables, including peppers. Early detection of this pathogen is crucial for a reduction in fungicide reliance and economic loss prevention. Traditionally, visual inspection has been a primary method for detection. However, symptoms often appear after the pathogen has begun to spread. This study employs the Deep Learning algorithm YOLO for single-class segmentation on plant images to extract spatial details of pepper leaves. The dataset included hyperspectral images at discrete wavelengths (460 nm, 540 nm, 640 nm, 775 nm, and 875 nm) from derived vegetation indices (CVI, GNDVI, NDVI, NPCI, and PSRI) and from RGB. At an Intersection over Union with a 0.5 threshold, the Mean Average Precision (mAP50) achieved by the leaf-segmentation solution YOLOv11-Small was 86.4%. The extracted leaf segments were processed by multiple Transformer models, each yielding a descriptor. These descriptors were combined in ensemble and classified into three distinct classes using a K-nearest neighbor, a Long Short-Term Memory (LSTM), and a ResNet solution. The Transformer models that comprised the best ensemble classifier were as follows: the Swin-L (P:4 × 4–W:12 × 12), the ViT-L (P:16 × 16), the VOLO (D:5), and the XCIT-L (L:24–P:16 × 16), with the LSTM-based classification solution on the RGB, CVI, GNDVI, NDVI, and PSRI image sets. The classifier achieved an overall accuracy of 87.42% with an F1-Score of 81.13%. The per-class F1-Scores for the three classes were 85.25%, 66.67%, and 78.26%, respectively. Moreover, for *B. cinerea* detection during the initial as well as quiescent stages of infection prior to symptom development, qPCR-based methods (RT-qPCR) were used for quantification of *in planta* fungal biomass and integrated with the findings from the AI approach to offer a comprehensive strategy. The study demonstrates early and accurate detection of *B. cinerea* on pepper plants by combining segmentation techniques with Transformer model descriptors, ensembled for classification. This approach marks a significant step forward in the detection and management of crop diseases, highlighting the potential to integrate such methods into *in situ* systems like mobile apps or robots.

Keywords: deep learning; segmentation; descriptor classification; image classification; vision transformers; *Botrytis cinerea*; gray mold; RT-qPCR; fungal biomass; precision agriculture

Academic Editor: Feng Yang

Received: 4 December 2024
Revised: 30 December 2024
Accepted: 10 January 2025
Published: 13 January 2025

Citation: Kapetas, D.; Kalogeropoulou, E.; Christakakis, P.; Klaridopoulos, C.; Pechlivani, E.M. Comparative Evaluation of AI-Based Multi-Spectral Imaging and PCR-Based Assays for Early Detection of *Botrytis cinerea* Infection on Pepper Plants. *Agriculture* **2025**, *15*, 164. https://doi.org/10.3390/agriculture15020164

1. Introduction

The demand for efficient farming practices in the horticulture and food industries is growing rapidly. The success of horticulture relies heavily on both the quality and quantity of its products. Fungal diseases can affect various plant parts, such as stems, leaves, flowers, and fruits, leading to reduced yields and lower overall crop production, which results in economic losses. Gray mold, caused by *Botrytis cinerea* (teleomorph *Botryotinia fuckeliana*), is one of the most destructive diseases globally. The pathogen infects over 1400 plant species, including economically important crops such as grapes, greenhouse vegetables, ornamental plants, and various fruits [1,2]. Efforts to develop eco-friendly control methods, such as RNA interference, biological control, and plant resistance inducers, are ongoing but have not yet matched the effectiveness of fungicides, which remain the primary management strategy. However, the success of chemical control is increasingly challenged by the development of fungicide resistance in *B. cinerea* populations, leading to the reduced or complete loss of control efficacy. Resistance has emerged globally against all site-specific fungicides due to mutations in the genes encoding the target sites [3].

Currently, diagnosis of *B. cinerea* Pers. is mainly based on the visual assessment of disease symptoms, as well as the isolation and identification of the pathogen from infected plant tissues [4]. This process is not only time-consuming and labor-intensive but also requires expertise. Early detection is crucial for enabling timely and effective management, which helps limit disease spread, reduce yield loss, and minimize chemical use [5]. Investigating a pathogen's gene expression patterns (expression profiles) at different stages is essential for understanding its behavior and progression. In recent times, the introduction of real-time polymerase chain reaction (qPCR) techniques to detect damaging pathogens has greatly improved quantitative analysis, combining the sensitivity of conventional PCR with the generation of a specific fluorescence signal only when the probe forms a stable hybrid with the complementary sequence of the amplicon. Reverse transcription quantitative PCR (RT-qPCR) is highly regarded as a benchmark for the high-precision, high-sensitivity, and fast measurement of gene expression. RT-qPCR has been extensively utilized for the detection of pathogens, including bacteria, viruses, and fungi, in plant tissues and soil [6,7]. The method is particularly advantageous for identifying latent infections where the pathogen load may be low and conventional detection methods may fail. By quantifying pathogen-specific mRNA, researchers can determine the presence and activity of pathogens in host tissues, which is crucial for effective disease management. Additionally, researchers can track the dynamic gene expression changes that occur as fungi colonize their hosts, adapt to environmental conditions, and respond to plant defense mechanisms, offering a clearer understanding of fungal–host interactions [8]. These studies are critical for advancing our knowledge of fungal pathogenicity and potential control strategies.

Recent advancements in computer vision utilizing deep learning (DL), have increasingly proven to be a valuable approach for detecting crop diseases [1]. When combined with hyperspectral imaging, detection performance shows improvement, with classification methods including decision tree-based classifiers, the K-nearest neighbor (KNN) algorithm, Support Vector Machine (SVM) models, Convolutional Neural Networks (CNNs), etc. [1]. Another emerging approach in this domain is the use of ensemble methods, which aim to improve performance in comparison to singular models. These methods encompass a variety of structural configurations and variations, making them a dynamic area of study [9].

The Transformer methodology was initially proposed by Vaswani et al. [10] in 2017 for Natural Language Processing tasks. Later, in 2020, it was also proposed as an applicable solution model for computer vision tasks by Dosovitskiy et al. [11], going by the name

Vision Transformer (ViT). This model demonstrated the ability to outperform the previously dominant Convolutional Neural Networks (CNNs) when provided with sufficiently large datasets. Of course, this model still had some drawbacks, like the quadratic compute-cost growth with the image resolution. Subsequent studies resulted in other transformer permutations based on the ViT, like the following: (a) the Swin Transformer model [12], with a focus on handling the computational complexity of ViT and the fixed scale; (b) the DeiT model [13], with a similar aim and an approach that involved utilizing a down-sampled pretrained ViT; (c) the VOLO model [14], targeting finer-level feature and context encoding in contrast with the coarse-level global dependency modeling performed by ViT; (d) the XCiT model [15], which applies a cross-covariance attention layer, granting a linear complexity to a task of quadratic complexity in the ViT model; and (e) the MaxViT model [16], which applies linear scaling in relation to image resolution, while also attempting to enable a global view of the entire network by the model.

In a recent study, Giakoumoglou et al. (2024) [17], approached the early detection of gray mold in cucumber plants. They utilized five multispectral images for each capture, combined with image-segmentation techniques using CNNs and ViT encoders. This method achieved an overall accuracy of 90.1% with a Dice Coefficient (DC) of 67.7% at an Intersection over Union (IoU) of 0.656. Subsequently, Christakakis et al. (2024) [18] focused on addressing dataset imbalance through the implementation of the augmentation technique Cut-and-Paste, enabling their model to manage an overall accuracy of 92% with a DC of 79.2% at the higher IoU of 0.816. Additionally, Scarboro et al. (2021) [19] approached *B. cinerea* detection in lettuce plants, using bispectral imaging as the base for single-pixel classification. Under laboratory conditions, their model performed with a True Positive rate of 95.25% while the False Positive rate was 9.316%. Qasrawi et al. (2021) [20] attempted the detection and classification of five tomato diseases, including *B. cinerea*, using smartphone imaging with a dataset of 3000 images. The models employed included neural networks, logistic regression, and clustering techniques, achieving 70.3%, 68.9%, and 70% classification accuracies, respectively. Regarding image classification, Lorente et al. (2021) [21] approached the task by applying standardized methods for descriptor extraction such as SIFT and SURF [22]. Then, they applied Bag of Visual Words (BoVW) [23] classification through an array of algorithms, and, based on the model and configuration, they reached accuracies ranging from 0.6 to 0.96. Nagasubramanian et al. (2019) [24] effectively applied hyperspectral imaging and a deep CNN architecture to classify soybean stems infected by *Macrophomina phaseolina*, a fungus that causes charcoal rot, an economically important disease of soybean worldwide. The model reached an accuracy of 95.73% with an 87% F1-Score regarding the infected class. This study showed that the near-infrared (NIR) wavelengths were most critical for detecting infected samples, while the physiological relevance of spectral regions was demonstrated through tests with wavelengths in the visible spectrum (400–700 nm) for disease detection. Similarly, Nguyen et al. (2021) [25] combined DL with hyperspectral imaging to enable the early detection of grapevine vein clearing virus (GVCV). The study demonstrated efficacy in differentiating healthy grapevines from those infected with GVCV, even in the early asymptomatic stages. Key wavelengths regions, including the NIR region (900–940 nm) and the visible spectrum (449–461 nm), were identified as critical for differentiation. Physiological changes, including decreased leaf water availability and a decline in the leaf pigment content, influenced reflectance patterns and proved integral to disease detection. These studies highlight the capability of hyperspectral imaging to enable the early detection of plant pathogens, and align with our approach to detect *B. cinerea*.

In our previous study, Kapetas et al. (2024) [26], an ensemble classification of descriptors through two approaches (KNN and LSTM) was effectively applied to classify

tomato leaves for the early detection of *B. cinerea.* This method involved the use of five hyperspectral images and one RGB image for each capture. The main goal of the current study is to build upon and enhance the previous methodology. The focus has shifted from tomato plants to pepper plants, with the main objective of this study still remaining the effective detection of *B. cinerea*, particularly during the early stages of infection or in latent infections where symptoms are not yet visible. A combination of molecular methods (RT-qPCR) for estimating the biomass accumulation of *B. cinerea* in host tissues, analyzing the expression of genes involved in fungal adaptation such as plant–pathogen interactions and pathogenicity, and computer vision techniques was applied. An imagery dataset was captured, consisting of leaf images covering all three classes to be distinguished: healthy, invisible-botrytis, and visible-botrytis. Each capture consisted of eleven images: five captured on hyperspectral wavelengths (460, 540, 640, 775, and 875 nm), one in RGB, and five from derived vegetation indices (CVI, GNDVI, NDVI, NPCI, NPCI). In comparison to our previous study, the introduction of the vegetation indices was an important step towards a more robust classification approach. YOLO [27] was utilized to segment and extract each leaf from these images. Subsequently, a descriptor was generated for each leaf for each of the 11 images of each capture using a suite of Transformer models. Also, a further preprocessing step that augmented the leaf-image quantity was introduced before the descriptor extraction to enhance the approach of our previous research. Finally, three descriptor classification solutions were applied, a KNN [28] algorithm, an LSTM-based (Long Short-Term Memory) [29] model, and a ResNet-based [30] model, to assign each leaf into one of the three distinct classes. As a further step towards improving the performance of the approach for classification and to ensure robustness, the multiple results obtained for each capture (one for each of the eleven images for each of the Transformer models) were ensembled via a weighted voting algorithm. This ensembling method effectively combined multiple image-classification predictions to yield the most accurate result.

2. Materials and Methods

A high-level flowchart is presented in Figure 1, outlining the key stages of the proposed methodology, which are discussed in detail in the following sections.

2.1. Biological Material and Inoculation Protocol

The bioassays were conducted following a completely randomized design in two controlled-environment plant growth chambers at Benaki Phytopathological Institute premises; one chamber housed *B. cinerea*-inoculated plants while the other contained mock-inoculated (non-inoculated) plants. Both chambers were maintained under the same environmental conditions (21 $\pm$ 1 °C, 16 h of light with a photosynthetic photon flux of 352.81 μmol/s, and 85–90% relative humidity).

The *B. cinerea* used in the study was isolated from naturally infected cucumber plants. For conidial suspension (10^5 conidia ml^{-1}) preparation, conidia were harvested by scraping the surface of 12-day-old culture gently and were suspended in sterile distilled water (SDW) supplemented with 2% sucrose and 0.01% Tween® 20 (Sigma-Aldrich, St. Louis, MO, USA) [17].

Pepper (*Capsicum annuum* L. cv. 'Soroksari') cv. KAPTUR F1 plants (Plantas S.A., Thiva, Greece) were artificially inoculated with the pathogen at the stage of four fully expanded leaves. Specifically, the second true leaf was inoculated by spraying the adaxial surface with the conidial suspension until it reached the point of run-off [17]. For mock-inoculated plants, the same method was carried out with the only difference being the omission of *B. cinerea* inoculum.

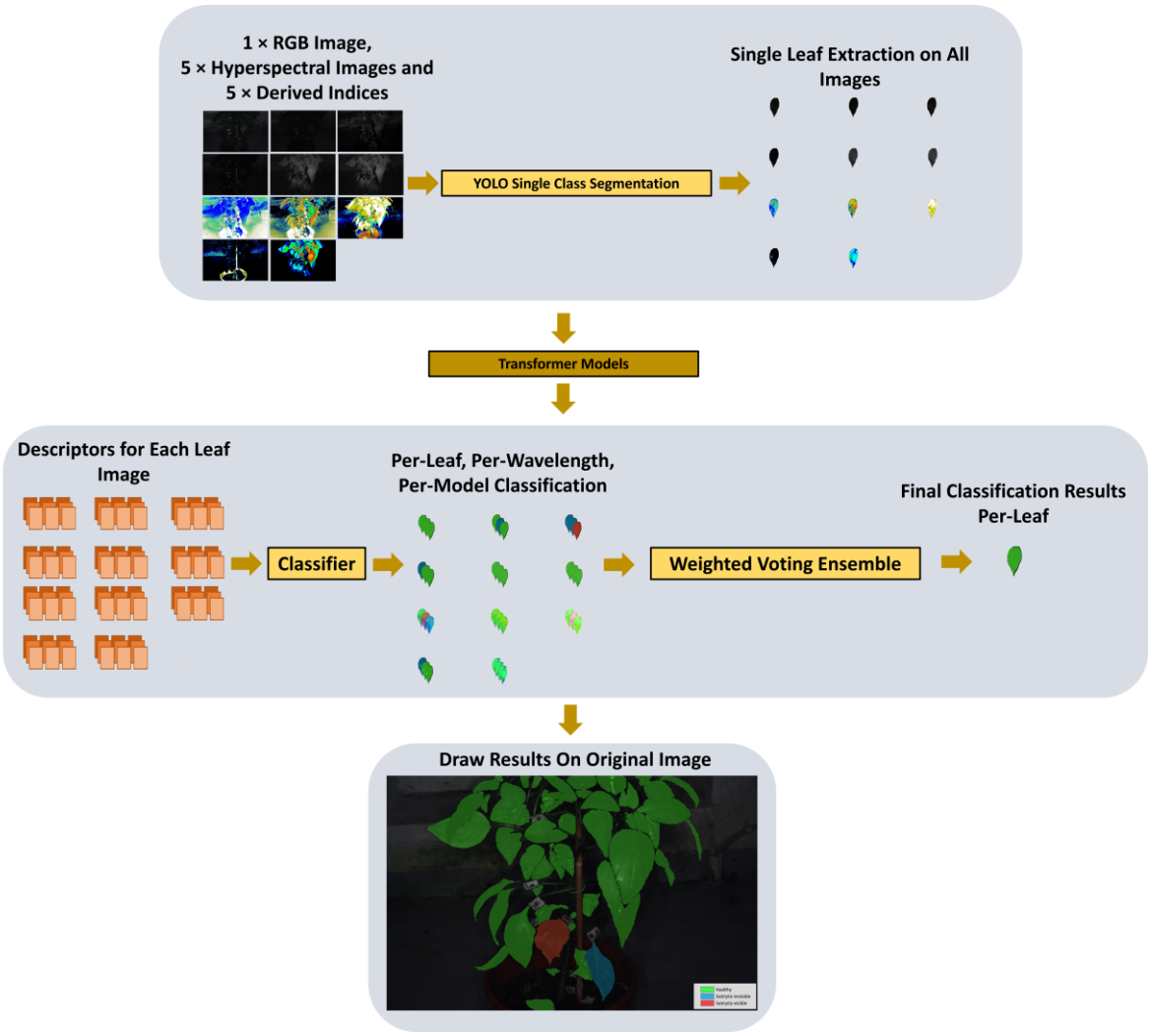

Figure 1. A high-level overview of the proposed methodology. The process begins with feeding the RGB image, five hyperspectral images, and five derived indices into the YOLO model to perform single-leaf segmentation across all images. The segmented leaf images are then processed through Transformer models to generate descriptors for each leaf. These descriptors are subsequently classified on a per-leaf, per-wavelength, and per-model basis. A weighted voting ensemble integrates these classifications to determine the final class for each leaf. Finally, the classifications are visualized on the original image, providing a comprehensive overview of the results. The leaves are categorized into three classes: green (healthy), blue (botrytis-invisible), and red (botrytis-visible).

2.2. In Planta Assays for Disease Intensity and Image Acquisition

The bioassays consisted of two sets of pepper plants: six plants inoculated with the *B. cinerea* (labeled 1–6) and three control plants that received a mock inoculation (labeled 1–3). The severity of disease was assessed twice weekly, based on the percentage of leaf area showing visible symptoms, and continued until the plants reached the end of their growth cycle. For each leaf, disease severity was assessed over a period of 65 days post-inoculation (dpi). The data were analyzed using a logit transformation to linearize disease progression over time [17]. Linear regression was then applied to determine the rate of disease progression and estimate the onset time of infection for each leaf.

Disease ratings were plotted over time to construct disease progress curves. The area under the disease progress curve (AUDPC) was calculated using the trapezoidal integration method [31,32]. Disease severity was expressed as a percentage of the maximum possible AUDPC for the whole period of the experiment, which is referred to as the relative AUDPC.

Initial and advanced symptoms of *B. cinerea* were captured using multispectral imaging. The images were subsequently captured at 1, 2, 5, 13, 20, 27, 34, 41, 48, 55, and 62 dpi, and included detailed views of individual leaves as well as the entire plants, showcasing all their foliage. The total number of captures obtained from *B. cinerea*-infected and mock-inoculated plants was 413 and 372, respectively. For imaging, a customized Qcell Phenocheck camera, Chania, Greece [33] was utilized, capturing six images from the visible and NIR spectrum (five images at the wavelengths 460 nm, 540 nm, 640 nm, 775 nm, and 875 nm and one RGB image), and generating the images of the derived indices (CVI, GNDVI, NDVI, NPCI, NPCI). The captured images had a resolution of 3000 × 1900 pixels. An example of the individual 11 images that comprised a capture of a *B. cinerea*-inoculated plant is shown in Figure 2.

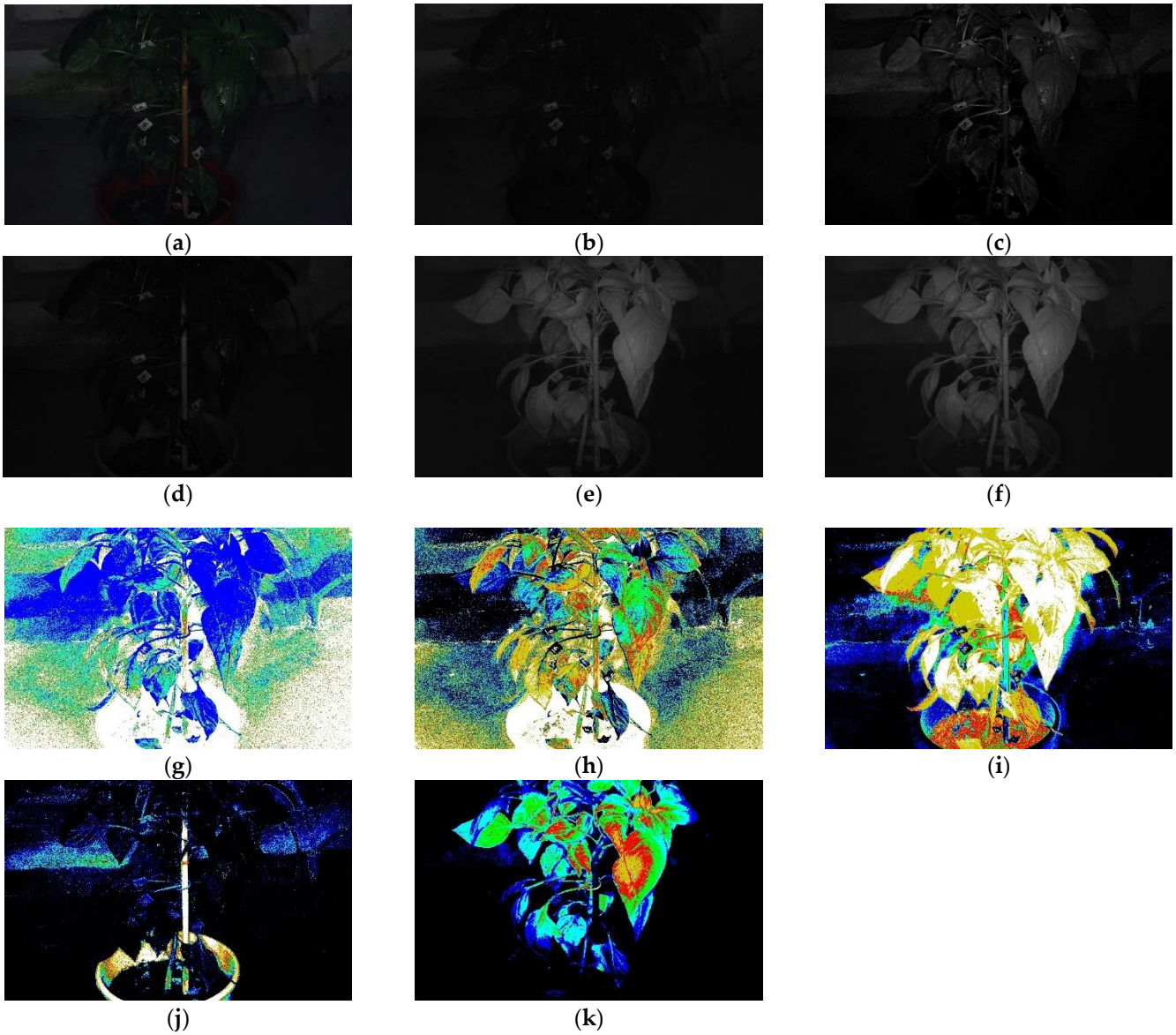

Figure 2. The eleven images that comprised a single capture: (**a**) RGB, (**b**) 460 nm, (**c**) 540 nm, (**d**) 640 nm, (**e**) 775 nm, (**f**) 875 nm, (**g**) CVI, (**h**) GNDVI, (**i**) NDVI, (**j**) NPCI, (**k**) PSRI.

2.3. Bioassays for Gene Expression Analyses and Image Acquisition at Early Stages of Infection

For gene expression analyses, the bioassays consisted of three plants (biological replicates) per treatment (mock-inoculated and *B. cinerea*-inoculated) and per time-point (1, 2 and 5 dpi). The second true leaf of each plant was sprayed with the conidial suspension as

described previously. At each time point, the mock- and *B. cinerea*-inoculated leaves were collected from each pepper plant, immediately used for multispectral imaging, flash frozen, ground in liquid nitrogen, and stored at −80 °C.

Total RNA was extracted using the NucleoSpin® RNA Plant and Fungi Kit (Macherey-Nagel GmbH & Co., Dueren, Germany) following the manufacturer's protocol. The concentration and purity of the RNA were assessed using a NanoDrop® ND-1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, NC, USA), and its integrity was confirmed through agarose gel electrophoresis [31,32].

cDNA synthesis, along with the removal of any residual genomic DNA, was carried out using the PrimeScript™ RT Reagent Kit with gDNA Eraser for Perfect Real Time (Takara Bio Inc., Kusatsu, Shiga Japan), following the manufacturer's protocol. Real-time PCR was performed using the StepOnePlus Real-time PCR system (Applied Biosystems, Thermo Fisher Scientific, Wilmington, NC, USA) with SYBR green (KAPA SYBR® FAST qPCR Master Mix (2X), KAPA Biosystems (Pty) Ltd., Cape Town, South Africa) as the fluorescent reporter. Each reaction mixture in 10 µL consisted of 5 µL of FAST qPCR Master Mix (2X), 0.2 µL of 10 µM of each gene-specific primer pair, 0.2 µL of 50X ROX High, and 1 µL of cDNA template.

Fungal biomass was quantified by assessing the relative expression of *B. cinerea* reference gene *BcRPL5* (*Bcin01g09620*) [34]. The relative expression of two defense-related genes in pepper, *Pathogenesis-related protein 1* (*PR1*) and *Defensin 1*, related to the salicylic acid (SA) and jasmonic acid (JA) pathways in pepper, respectively, were also estimated in both pathogen-inoculated and mock-inoculated samples [35]. Ubiquitin-conjugating protein CaUbi3 (Accession Number: AY486137.1) was used as an internal reference [36]. The relative expression level of the selected genes was calculated with the $2^{-\Delta\Delta CT}$ method [32]. The sequences of all primers utilized in this study are listed in Supplementary Table S1. All reactions were performed in duplicate. Additionally, melting curve analysis was performed to confirm the absence of non-specific products and primer dimers.

2.4. Image Annotation and Segmentation Techniques

Roboflow [37] was employed to annotate the 785 captures for a polygon encapsulation of each leaf's area in the images. According to the development of *B. cinerea* disease, each leaf was assigned a class. Leaves with 0.1–5% infected leaf area were labeled as "botrytis-invisible" while those with 5–100% infected leaf area were categorized as "botrytis-visible". All other leaves were labeled as "healthy." For simplicity and ease of reference, the classes were renamed as follows: "healthy" leaves as "Class 0", "botrytis-invisible" leaves as "Class 1" and "botrytis-infected" leaves as "Class 2". A depiction of the true annotations mask, overlaid over an RGB image, is depicted in Figure 3.

Image segmentation is the process of identifying and partitioning distinct regions within an image, where each segment represents a specific object or area of interest [38]. In this study, segmentation was performed using a DL model based on the YOLO architecture.

This model is capable of recognizing image features by learning generalizable object representations from image datasets. To achieve a balance between accuracy and inference speed, a pretrained YOLO-Small model was selected from the Ultralitics library [39] and fine-tuned. The fine-tuning process involved training models from the YOLOv8 and YOLOv11 architectures for 100 epochs using the SGD optimizer with a learning rate of 0.01. Fine-tuned models were tested at resolutions of 1024 × 1024 and 1600 × 1600, with batch sizes of 4 and 8.

Figure 3. RGB image overlaid with the ground truth annotations. The leaves are categorized into three classes: green (healthy), blue (botrytis-invisible), and red (botrytis-visible).

2.5. Data Processing and Augmentation Techniques

This study approached leaf classification by isolating each leaf on its own separate new image, given that each original image shows more than one leaves. The area on the image that was not occupied by that one leaf was fully transparent (dataset-a). To facilitate compatibility with the Transformer models outlined in Section 2.6, the individual leaf images were resized to a uniform dimension of 384 × 384 pixels. Down-sampling from the original resolution of 3200 × 1900 pixels to the resolution of 384 × 384 pixels led to a severe loss of information for the single leaf, particularly for leaves occupying a small part of the original image. Additionally, the resulting images contained a large number of transparent pixels that essentially provided almost no information. To address these issues, two zooming techniques were employed. The first method effectively represents a four-times zoom (dataset-b), which is achieved by cropping each leaf image to half its original width and height before resizing. This approach aimed to preserve finer details while ensuring even the largest leaves of the dataset were still fully visible within the images. The second method (dataset-c) focused on cropping the images closely around each leaf's bounding box area which maximized the preserved leaf-pixel information to the final image but, as each leaf fully occupied the available image space, all leaf-size context was lost. Figure 4 illustrates how datasets-a, -b, and -c were derived from the original images.

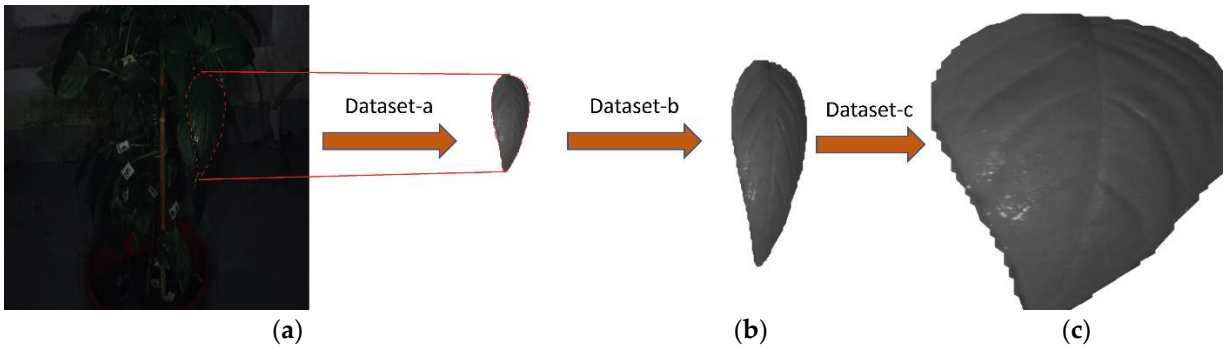

Figure 4. A depiction of an example image of a leaf from datasets (**a**): as extracted directly from the original image; (**b**): with a four-times zoom; (**c**): with a crop around the leaf area. Additionally, (**a**) portrays how the leaf image was extracted from the original image in a visual manner.

The resulting dataset from extracting single-leaf images contained 107,481 images in the "healthy" class, 473 in the "botrytis-invisible" class, and 1309 in the "botrytis-visible" class. This dataset included eleven images per capture (five from hyperspectral wavelengths, one RGB image, and five from derived indices) and included only leaves occupying a minimum of 0.2% of the image area, as smaller leaves provided insufficient pixel quantity to contain meaningful information for analysis by the models.

Since the classification approach (outlined in Section 2.7) required a relatively balanced count of images in each class, the disproportionately large number of "healthy" class images was reduced. During this reduction, a key point was that at least one capture from each day of the experiment remained in the "healthy" class, preserving all eleven images per capture. After reduction, the "healthy" class contained 1507 images, while the other two classes remained unchanged.

To further enhance model performance, additional preprocessing was applied. Specifically, before reducing the "healthy" class, data augmentation was performed to expand the "botrytis-invisible" and "botrytis-visible" classes by increasing their image count. This augmentation involved the randomized scaling and rotation of individual leaves to create realistic variations, while further common augmentations such as blurring, brightness, exposure, or other color-based modifications were rejected as they would undermine the objective of using separate wavelengths where the coloring is an important part of the information. Augmentation factors of 3.7 and 2 were applied to the "botrytis-invisible" class and the "botrytis-visible" class, respectively. Although these factors may seem high, they were necessary due to the low original image counts, especially in the "botrytis-invisible" class, which had only 43 separate captures. Then, applying an appropriate reduction to the images of the "healthy" class, like described before, the final counts of the dataset were 2882, 1749, and 2575 images for each of the classes.

Finally, the dataset was divided into training and testing subsets leading to 1133, 352, and 1023 training images and 374, 121, and 286 testing images for each class, respectively, for the original dataset with the reduced "healthy" images, while for the dataset where the data augmentation was applied, the split led to 2266, 1320, and 1957 training images and 616, 143, and 407 testing images. A key criterion during the splitting process was that at least one capture for each day and for each class was included in both the training and testing sets, with all 11 images per capture to be assigned together to either set.

2.6. Transformer Models for Descriptor Extraction

A machine learning model's layer produces a descriptor as output, which is a vector formed to characterize the input data, which, in this case, is an image, while a global descriptor is the output of the final layer before the model's output layer [40]. In this study, an array of 9 Transformer models was utilized to extract 9 global descriptors for each of the eleven images of each capture. More specifically, the timm [41] library's models were employed, which are pretrained on ImageNet-1K [42]. The nine transformer models were as follows: ViT-L (P:16 × 16), ViT-L (P:32 × 32), ViT-B (P:16 × 16–C), ViT-B (P:32 × 32–C), Swin-L (P:4 × 4–W:12 × 12), VOLO (D:5), XCIT-L (L:24–P16 × 16), DEIT-L (P16 × 16), and MaxViT-L.

2.7. Classification Methods

In Section 2.6, it is described that a different descriptor was extracted on a per-image basis for the images that refer to specific singular leaves. Since there were 11 descriptors for each leaf image, 11 classification actions would be performed independently. Three different approaches were employed in order to classify these extracted descriptors.

KNN classification was the first selected approach. In this method, a separate BoVW was created for each of the eleven images forming a capture. Each BoVW served as a library of visual features for all descriptors from the training set. Then, to classify descriptors in the testing set, using a distance metric, each descriptor was sequentially compared to all items in the corresponding BoVW for the same wavelength. The distances obtained were listed and the classification was achieved averaging the K smallest distances. This study used three different distance metrics: Euclidean, Bray–Curtis, and Cosine.

The second method involved an LSTM model, built using Keras [43]. In this setup, an LSTM network was trained on the descriptors to capture significant features, using an LSTM layer, a Batch Normalization layer, and then another LSTM layer. A dense layer with a linear activation function mapped the output to the corresponding image label. The LSTM model was trained on descriptors extracted from one of the Transformers for all training images for each wavelength, and separate LSTM models were trained for each Transformer model. This led to the creation of 9 LSTM models in total, one for each of the Transformer models. The learning rate of the models was set to 0.001 initially, with a learning rate scheduler in place. Training spanned 12 epochs for each wavelength, and the batch size was set to 128.

The third solution was to use a ResNet architecture, developed again with Keras. The ResNet model consisted of several residual blocks, each configured with Conv1D layers, Batch Normalization, and ReLU activations to refine the feature extraction process. Each residual block included a skip connection to mitigate the vanishing gradient problem, allowing gradients to flow more effectively during backpropagation. The model began with an initial Conv1D layer with 512 filters and a kernel size of 3, followed by a sequence of residual blocks with varying filter sizes and strides to capture hierarchical features. Dropout layers were integrated between blocks to improve generalization and reduce overfitting. The feature representations were then compressed using Global Average Pooling, and the final dense layer with a softmax activation function mapped these features to the 3 class labels. Similarly to the LSTM models, a separate ResNet model was trained for each descriptor set coming from each of the 9 Transformer models, while each ResNet model was trained on descriptors from all 11 images per capture. Again, like in the LSTM approach, the initial learning rate was initialized at 0.001, with a learning rate scheduler in place. The model was trained for 12 epochs for each wavelength.

To tackle the class imbalance issue due to the different numbers of images per class, a Per-Class Confidence Multiplier (PCCM) was used on the classification outcomes. The PCCM was calculated by taking the inverse of the number of images in each class and multiplying it by the number of images in the class with the most images. Equation (1) illustrates the PCCM calculation, while Table 1 provides the resulting PCCM values.

$$PCCM_i = \frac{N_{max}}{N_i},\ i \in [0,2] \quad (1)$$

where N_i represents the number of training images in class i and N_{max} represents the number of training images in the class with the most images.

Table 1. Confidence multiplier per class.

Class	Dataset Without Data Augmentation		Dataset with Data Augmentation	
	Training Images	Confidence Multiplier	Training Images	Confidence Multiplier
0—healthy	1133	1	2266	1
1—botrytis-invisible	352	3.25	1320	1.716
2—botrytis-visible	1023	1.118	1957	1.157

2.8. Evaluation Metrics

For the evaluation of the results of the classification approaches utilized in this study that differentiated images into three classes, an important note is that the results classified each image into one and only one class. The metrics that were derived from those results were True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

The performance metrics for the models—accuracy, precision, recall, and F1-Score [44]—were calculated. The formulas for these metrics are as follows:

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

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

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

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The TP, FP, TN, and FN metrics were calculated for each descriptor in the testing set for all three classification approaches (KNN, LSTM, ResNet) by comparing the predictions with the ground truth class. The derived performance metrics were calculated on all abstraction levels, including per-class per-day, per-class for all days, and per-day for all classes.

2.9. Classification Ensemble

As described in Section 2.6, each of the 11 images of a capture had its own descriptor for each of the 9 Transformer models. During classification, a different prediction was made for each descriptor from a single capture. However, the final output should be one and only one prediction for the class of each capture. Therefore, a method was necessary to consolidate these multiple predictions into one definitive result. That method should offer higher classification performance compared to any one standalone descriptor set from a single image set and Transformer model.

To achieve that, an ensemble approach was employed, aggregating the multiple predictions relevant to each capture and producing the final result. The calculations for the ensemble were stored separately for multiple wavelengths and multiple model combinations. This structure enabled the possibility of calculating the ensemble results of multiple ensemble calculations for producing the final outcome.

To calculate the ensemble result of multiple predictions, the formula in Equation (2) was used to determine the confidence of each class as the weighted sum of predictions for that class, where each prediction was multiplied with that class's F1-score for that image set (out of the 11) of the corresponding descriptor set (out of the 9). Then, the class with the highest confidence among the three was declared as the selected option. To determine which combination of descriptor sets and image sets yielded the best final prediction results, this process was iteratively applied to all combinations of the 9 descriptor sets, meaning that not all sets were required to achieve the best result, and the method was also applied across multiple multi-image-set combinations.

$$C_i = \sum_{k=1}^{W} \sum_{j=1}^{M} \left(p_{j,k} \cdot A_{j,k} \right) \quad (2)$$

where C_i corresponds to the confidence score for class i across all models and wavelengths, $p_{j,k}$ indicates whether the prediction of model j for wavelength k corresponds to class i (1 if it does, 0 otherwise), $A_{j,k}$ refers to the F1-Score of model j for wavelength k, M represents

the total number of models, and *W* represents the total number of images per capture that are employed.

3. Results and Discussions

3.1. Assessment of Gray Mold Severity in Pepper Plants

Disease severity was evaluated as the percentage of leaf area showing gray mold symptoms for the total leaf area per leaf of all nine leaves per pepper plant, and was periodically recorded over 65 dpi. Disease progress was monitored at intervals of 3 to 5 days.

Concerning the second artificially inoculated leaf, disease symptoms were observed only in two out of six pepper plants used in the bioassay at 9 and 16 dpi, while disease progress was recorded only in one of them (plant No. 4) (Figure 5). The other inoculated leaves were naturally infected later, with the exception of one leaf that remained uninfected (Figure 5).

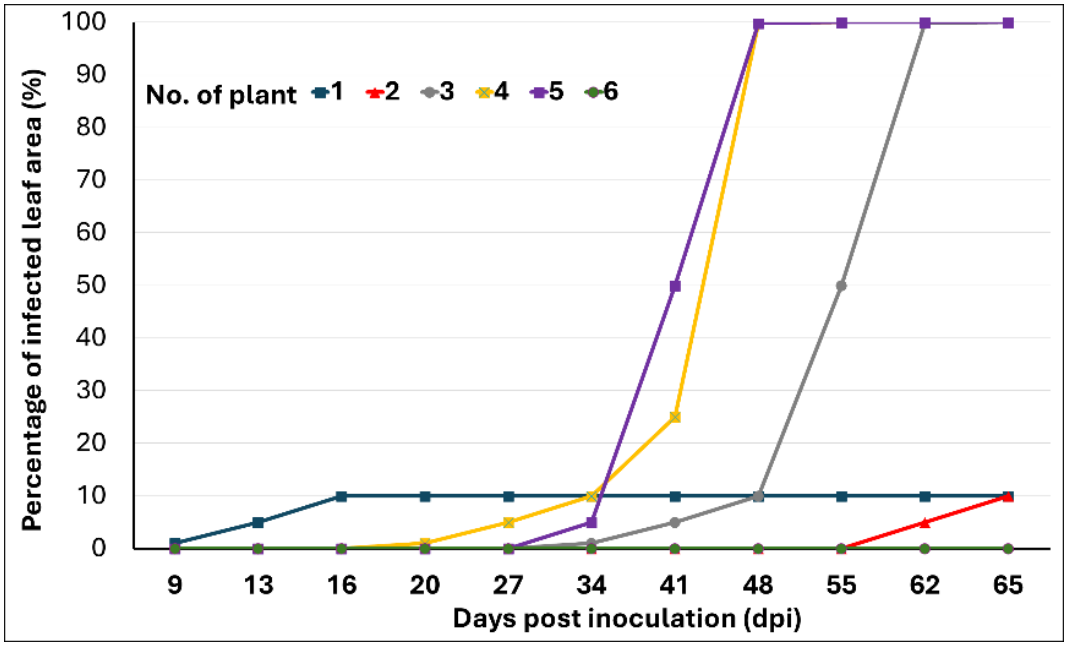

Figure 5. Disease severity (% of leaf area showing gray mold symptoms) as recorded over a period of 65 dpi on the 2nd artificially inoculated leaf of six plants.

The disease severity in the pepper leaves progressed at a slow rate, with the AUDPC reaching 17% for the second artificially inoculated leaf. Additionally, the remaining leaves, which were naturally and accidentally infected by *B. cinerea*, showed AUDPC values ranging from 4% to 13% (Figure 6). On the contrary, flowers and fruits were more susceptible to the pathogen, displaying severe symptoms such as rot, with their tissues covered by dense gray masses of conidia and mycelia (Figure 7D,E). Quiescent infections in asymptomatic leaves are critical, as they may transition to an aggressive state as the leaves age and enter senescence [45].

The progression of the disease was influenced by a combination of factors, such as leaf developmental stage, spatial arrangement on the stems, nutritional composition of the leaf tissue, plant defense mechanisms against the invading pathogen, and secondary disease cycles initiated by airborne *B. cinerea* conidia released not only from the initial fungal infections on the artificially inoculated leaf tissues but also from other more susceptible plant tissues, such as flowers and fruits (Figure 7). The pathogen is most destructive on mature or senescent host tissues, but when it invades these tissues at an earlier stage of

crop development it remains quiescent for extended periods and becomes active once environmental conditions become favorable and host physiology changes [46]. The occurrence of latency or quiescent infections represents an important component of the disease cycle and complicates disease management efforts.

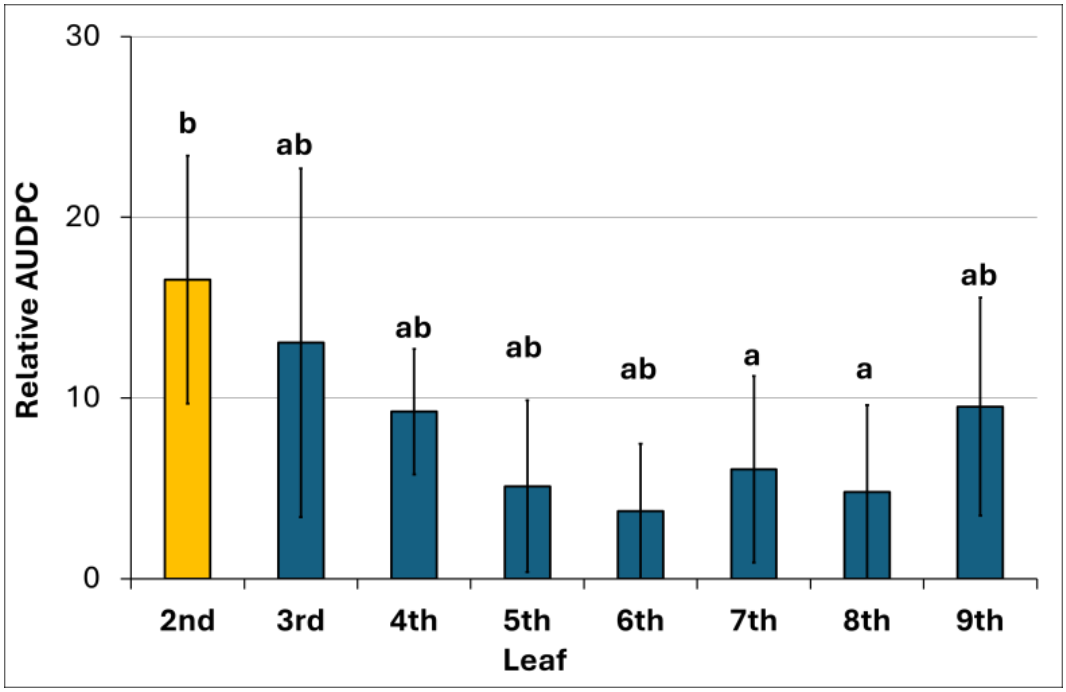

Figure 6. Disease severity expressed as the percentage of the maximum possible area under disease progress curve (AUDPC) for the whole period of the experiment (65 dpi) on the 2nd artificially inoculated leaf (yellow column) and on the 3rd and up to the 9th naturally and accidentally infected leaves (blue columns) of the six plants used in the bioassay. Bars indicate standard deviation. Columns with the same letter are not different according to Fisher's LSD multiple range test ($p \leq 0.05$).

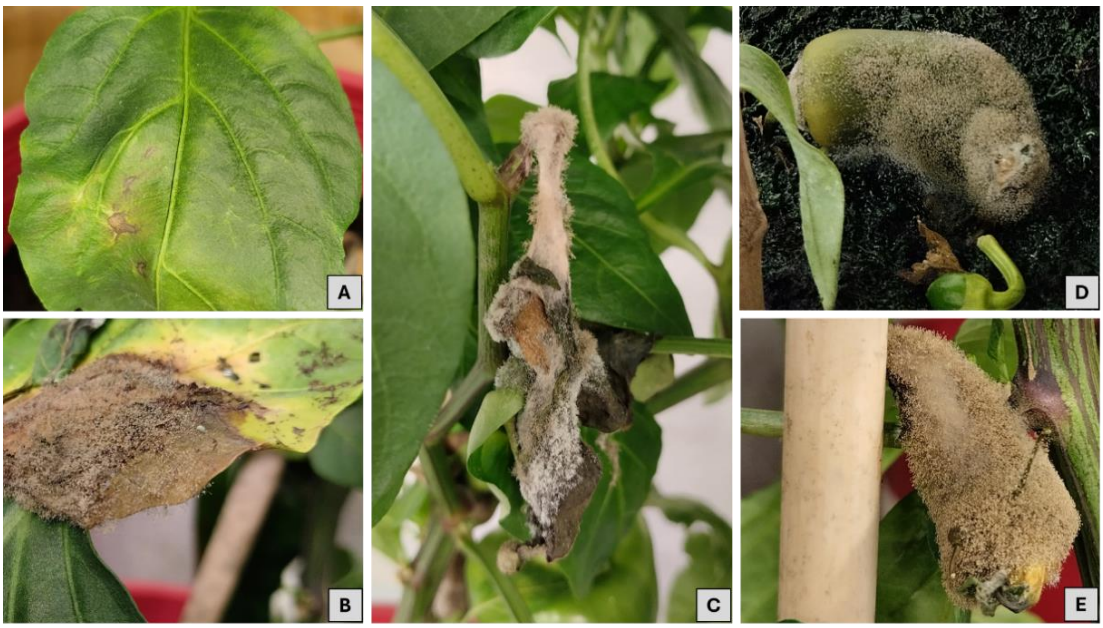

Figure 7. Gray mold symptoms: (**A**) Early and (**B**,**C**) late stages of leaf infection. (**D**,**E**) Flower and fruit infection.

3.2. Pepper Response to Botrytis cinerea Infection

Due to the challenge of the visual detection of gray mold symptoms in the initial stages of infection, fungal progression was evaluated by the relative expression of the fungal reference gene *BcRPL5* (*Bcin01g09620*). Specifically, the biomass of *B. cinerea* in the plant tissues of mock- and *B. cinerea*-inoculated second pepper leaves was quantified through the relative transcript abundance of *BcRPL5* via RT-qPCR. The results showed that the inoculated leaves exhibited a significant increase in fungal biomass within their tissues compared to mock-inoculated ones at all time points (1, 2, and 5 dpi) (Figure 8). The relative transcript levels of the *BcRPL5* gene increased significantly day by day, reaching their highest levels at 5 dpi (Figure 8). This gradual increase in the relative transcript levels of the fungal reference gene indicates that *B. cinerea* was actively expressing its genes over time, reflecting the pathogen's growth and pathogenicity, particularly during the early stages of infection when no visible symptoms were present.

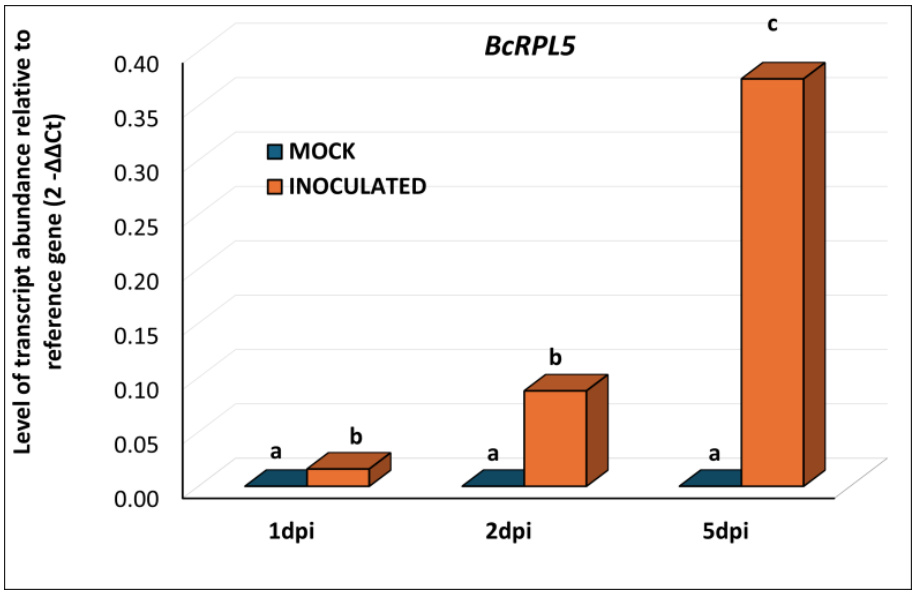

Figure 8. The relative *Botrytis cinerea* fungal abundance determined by RT-qPCR analysis of the fungal reference gene *BcRPL5* (*Bcin01g09620*) in mock- and *Botrytis cinerea*-inoculated second pepper leaves at 1, 2, and 5 dpi. Columns represent the means of three independent leaf samples per treatment (mock- and *B. cinerea*-inoculated). Columns with different letters are statistically different according to Fisher's LSD multiple range test ($p \leq 0.05$).

Additionally, we conducted analyses of the expression of *Capsicum annuum* defense-related genes to understand the hosts' responses to *B. cinerea* infection. Specifically, real-time RT-PCR analyses were performed to investigate the relative expression of the jasmonic acid-dependent gene *DEF1*, encoding the *Capsicum annuum* defensin 1 protein, and the SA-dependent gene *PR1*, encoding the *Capsicum annuum* pathogenesis-related protein 1 [47]. Upon fungal infection, *DEF1* expression was downregulated in *B. cinerea*-inoculated leaves compared to mock-inoculated ones at 2 and 5 dpi (Figure 9, left diagram). On the contrary, *PR1* was overexpressed in inoculated leaves at 5 dpi (Figure 9, right diagram). The downregulation of *DEF1*, a marker associated with the JA signaling pathway, suggests that JA-mediated defense responses were suppressed in response to *B. cinerea* infection. This could indicate a strategy by the pathogen to evade the plant's defenses that are typically effective against necrotrophic pathogens. The overexpression of *PR1*, a hallmark of the SA signaling pathway, highlights a shift in the plant's defense strategy. This indicates that, although JA-mediated defenses may be compromised, the SA pathway is activated

to combat the fungal infection, suggesting a complex interplay between these two pathways. The contrasting regulation of *DEF1* and *PR1* supports the notion of antagonistic interactions between the SA and JA signaling pathways. While JA is generally associated with defense against herbivores and necrotrophic pathogens like *B. cinerea* [48,49], the activation of the SA pathway indicates a localized immune response to fungal infection. The observed alterations in gene expression at different time points (2 and 5 dpi) suggest that the plant's defense response is dynamic. The initial downregulation of *DEF1* might benefit the pathogen, but as the infection progresses, the activation of *PR1* indicates a robust activation of the SA pathway to manage the growing threat posed by the pathogen.

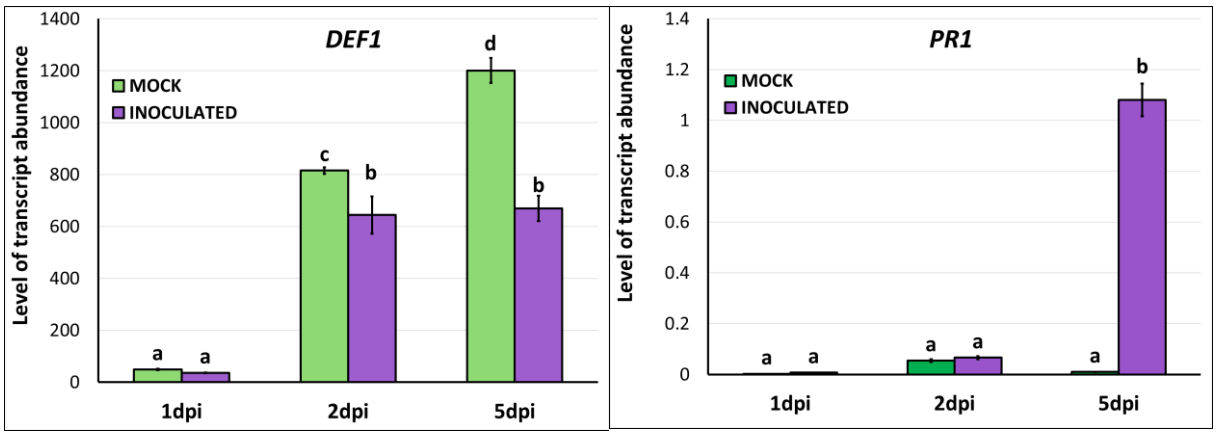

Figure 9. The relative expression of the disease-responsive pepper genes *DEF1* (*Defensin 1*) and *PR1* (*Pathogenesis-related protein 1*) in mock- and *Botrytis cinerea*-inoculated second pepper leaves at 1, 2, and 5 dpi via RT-qPCR. Columns represent the means of three independent leaf samples per treatment (mock- and *B. cinerea*-inoculated). Bars indicate standard deviation and columns with the same letter are not different according to Fisher's LSD multiple range test ($p \leq 0.05$).

These temporal changes in gene expression highlight a dynamic defense mechanism where the host adjusts its strategy based on the pathogen's progression. Thus, the observed variations in *B. cinerea* biomass in pepper tissues and the hosts' responses to the pathogen's attack align with the observed patterns of disease progression, illustrating the complexity of plant–pathogen interactions.

3.3. Performance Evaluation of Segmentation

From the YOLO-Small models that were trained, the best proved to be the YOLOv11-Small, trained with a resolution of 1600 × 1600 pixels and batch size set to four reaching an 86.4% mAP_{50}. Table 2 presents the metric results for the trained models. Also, Figure 10 provides visualized examples of the leaf segmentation in action.

Table 2. Performance metrics (mAP50, recall, and precision) for the YOLO models.

Model	Image Size	Batch Size	mAP_{50}	$mAP_{50\text{-}95}$	Recall	Precision
YOLOv8-Small	1024	4	0.859	0.611	0.76	0.849
YOLOv8-Small	1024	8	0.853	0.599	0.753	0.855
YOLOv8-Small	1600	4	0.860	0.624	0.767	0.865
YOLOv11-Small	1024	4	0.861	0.613	0.768	0.864
YOLOv11-Small	1024	8	0.857	0.601	0.756	0.859
YOLOv11-Small	1600	4	0.864	0.625	0.770	0.869

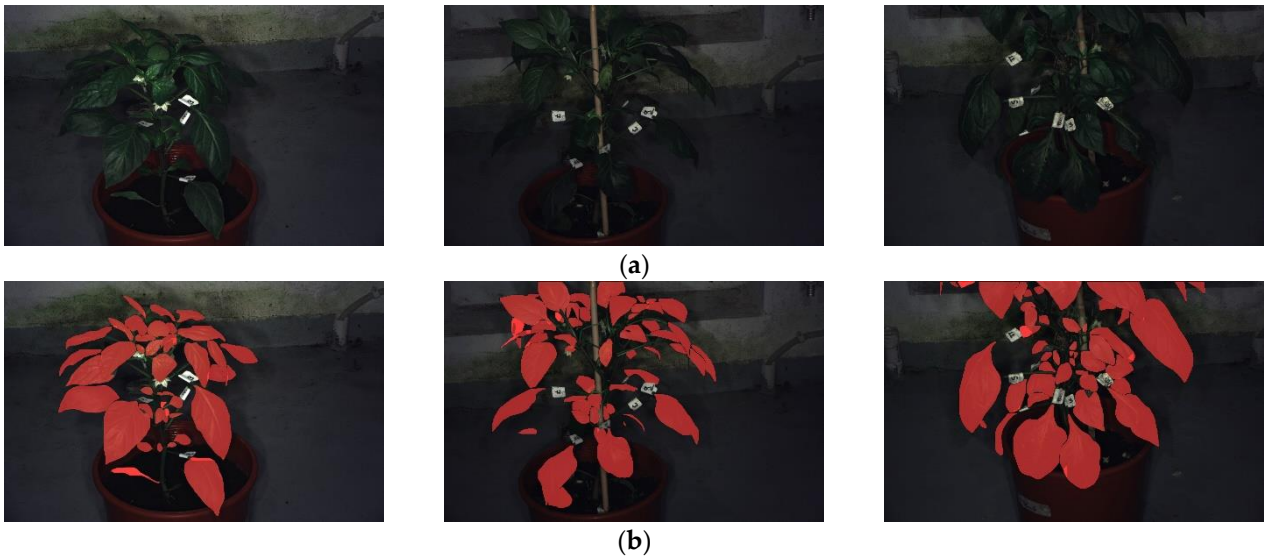

Figure 10. (**a**) RGB images captured by the camera and (**b**) corresponding YOLO predictions visualized in red.

3.4. Performance Evaluation of Classification

All three classification methods (KNN, LSTM, and ResNet) were applied to all three datasets (the original dataset-a, the 4× zoomed dataset-b, and the leaf-zoomed dataset-c), and all descriptors from all nine Transformer models were classified. Subsequently, an ensemble classification of the eleven separate descriptors for each capture was performed. The evaluation metrics included accuracy, precision, recall and F1-Score. The F1-Score was chosen as the decisive metric to evaluate the model's performance, since it provides a balanced measure by considering both precision and recall, whereas accuracy alone can sometimes give a misleading representation of a model's effectiveness. Table 3 shows in detail the F1-Scores (%) of each classification method for each of the three datasets for each descriptor set derived from the nine Transformer models across all classes. Based on that data, dataset-b outperformed dataset-a and dataset-c by 1.76% and 1.46% on average, respectively. This suggests that dataset-a was always outperformed by the zooming counterparts and that the leaf original shape and size information were important to be preserved in balance rather than maximizing pixel information, which was preserved in dataset-c. Based on that outcome, all subsequent experiments focused only on dataset-b, while the other two datasets were excluded. Additionally, it is worth mentioning that the KNN solution in this case was calculated using a K value of 3 and the Euclidian distance function.

Table 3. F1-Score (F1) percentage metrics on a per-Transformer model, per-solution (KNN/LSTM/ResNet), and per-dataset basis.

Models	(a) No-Zoom Dataset			(b) 4-Times-Zoom Dataset			(c) Leaf-Zoom Dataset		
	KNN	LSTM	ResNet	KNN	LSTM	ResNet	KNN	LSTM	ResNet
ViT-L (P:16 × 16)	71.01%	75.36%	56.52%	69.57%	71.83%	67.61%	68.20%	71.01%	66.28%
ViT-L (P:32 × 32)	66.67%	69.57%	49.28%	71.01%	69.01%	47.89%	75.61%	60.87%	46.94%
ViT-B (P:16 × 16–C)	68.12%	65.22%	63.77%	71.01%	70.42%	47.89%	68.20%	69.57%	46.94%
ViT-B (P:32 × 32–C)	69.57%	71.01%	49.28%	71.01%	70.42%	60.56%	72.65%	71.01%	59.37%
Swin-L (P:4 × 4–W:12 × 12)	65.22%	71.01%	68.12%	69.57%	69.01%	50.70%	65.23%	68.12%	49.71%
VOLO (D:5)	71.01%	60.87%	49.28%	71.01%	77.46%	49.30%	69.68%	73.91%	48.33%
XCIT-L (L:24–P16 × 16)	75.36%	69.57%	49.28%	65.22%	64.79%	61.97%	72.65%	69.57%	60.75%
DEIT-L (P16 × 16)	71.01%	60.87%	36.23%	68.12%	69.01%	47.89%	65.23%	71.01%	46.94%
MaxViT-L	68.12%	68.12%	49.28%	71.01%	67.61%	47.89%	63.75%	65.22%	46.94%

Table 4 shows the results of applying the PCCM. The data demonstrate an average of a 1.83% improvement compared to cases where the PCCM was not applied. Thus, any further experiments will always include the PCCM.

Table 4. F1-Score (F1) percentage metrics on a per-Transformer model, per-solution (KNN/LSTM/ResNet), and per-dataset basis, when applying the PCCM.

	(b) 4-Times-Zoom Dataset		
Models	**KNN**	**LSTM**	**ResNet**
ViT-L (P:16 × 16)	66.67%	70.42%	69.45%
ViT-L (P:32 × 32)	76.81%	69.01%	49.19%
ViT-B (P:16 × 16–C)	72.46%	67.61%	49.19%
ViT-B (P:32 × 32–C)	75.36%	67.61%	62.22%
Swin-L (P:4 × 4–W:12 × 12)	73.91%	70.42%	52.09%
VOLO (D:5)	68.12%	73.24%	50.64%
XCIT-L (L:24–P16 × 16)	68.12%	69.01%	63.66%
DEIT-L (P16 × 16)	71.01%	73.24%	49.19%
MaxViT-L	72.46%	70.42%	49.19%

Applying the data augmentations to dataset-b, a noticeable average improvement of 4.44%, 1.07%, and 24.3% was observed for each of the KNN, LSTM, and ResNet classification solutions, respectively. Table 5 shows the F1-Scores on the same level as the previous table with the data augmentations applied. Based on this performance improvement, all further experiments will have the data augmentation applied.

Table 5. F1-Score (F1) percentage metrics on a per-Transformer model, per-solution (KNN/LSTM/ResNet), and per-dataset basis, with the data augmentations applied.

	(b) 4-Times-Zoom Dataset		
Models	**KNN**	**LSTM**	**ResNet**
ViT-L (P:16 × 16)	73.58%	66.98%	69.81%
ViT-L (P:32 × 32)	72.64%	69.81%	66.04%
ViT-B (P:16 × 16–C)	75.47%	72.64%	66.98%
ViT-B (P:32 × 32–C)	77.36%	74.53%	69.81%
Swin-L (P:4 × 4–W:12 × 12)	75.47%	75.47%	58.49%
VOLO (D:5)	73.58%	72.64%	73.58%
XCIT-L (L:24–P16 × 16)	71.70%	70.75%	72.64%
DEIT-L (P16 × 16)	77.36%	69.81%	71.70%
MaxViT-L	76.42%	65.09%	66.04%

The KNN solution did not undergo training since it is an algorithmic solution. As explained in Section 2.7, a separate BoVW was created for each of the 11 descriptors corresponding to each capture from the 9 Transformer models. This modularity allowed classifications of different image sets with no residual effect from any other image set. The per-image-set, per-Transformer model F1-Scores (%) of the KNN solution are presented in Table 6.

On the other hand, for the LSTM and ResNet models, nine separate models were trained, one for each Transformer model. All descriptors from all the different image sets (11 up to this point) contributed to the model training and, thus, to the result. This approach introduces a challenge: if one or more image sets underperformed in comparison to the others during training, the performance of the whole model would have been brought down. Observing the average F1-Score values in Table 6 reveals noticeable performance variations across the 11 image sets.

Table 6. F1-Scores (%) of the KNN solution on a per-Transformer model, per-image-set basis.

Models	RGB	CVI	GNDVI	NDVI	NPCI	PSRI	460 nm	540 nm	640 nm	775 nm	875 nm
ViT-L (P:16 × 16)	58.49	67.30	63.84	64.15	61.01	62.89	52.52	57.23	55.35	62.58	58.49
ViT-L (P:32 × 32)	56.60	67.61	64.15	61.32	59.12	61.95	58.18	62.26	50.00	60.69	58.18
ViT-B (P:16 × 16–C)	64.15	63.52	59.12	64.47	61.64	64.47	57.23	66.35	59.12	59.43	62.26
ViT-B (P:32 × 32–C)	64.78	65.09	63.21	55.03	61.32	65.72	57.86	59.43	54.72	58.81	59.12
Swin-L (P:4 × 4–W:12 × 12)	59.12	64.78	56.92	62.26	61.64	68.24	55.97	58.81	55.03	59.43	60.69
VOLO (D:5)	61.01	63.52	69.81	61.01	62.26	59.43	50.63	58.49	55.66	54.72	52.52
XCIT-L (L:24–P16 × 16)	66.98	68.87	60.69	61.64	61.64	63.84	55.97	62.89	62.58	57.86	54.72
DEIT-L (P16 × 16)	63.84	67.92	64.15	53.46	57.86	55.35	54.09	60.06	61.32	59.43	61.01
MaxViT-L	62.26	59.12	60.06	65.72	59.75	65.41	56.60	67.61	56.92	55.97	57.55
Average Value Across All Models	61.91	65.30	62.44	61.01	60.69	63.03	55.45	61.46	56.74	58.77	58.28

To further improve the performance of the models, the selective removal of some of the worse-performing image sets from the training data was explored. An iterative process was employed that tested the removal of all combinations of image sets from the wavelengths 460, 640, 775, and 875 nm. For the LSTM approach, the best performing results were yielded by models where the image sets from the 460 and 640 nm wavelengths were discarded. This approach outperformed the previous solution by a per-model average F1-Score of 2.36%. For the ResNet approach, however, the case with all 11 image sets proved best. Subsequent analysis with the LSTM approach included neither the 460 nor the 640 nm wavelengths. Table 7 shows the updated F1-Scores (%) on a per-model basis for the LSTM models trained without the image sets of the 460 or the 640 nm wavelengths.

Table 7. The per-model F1-Scores (F1) percentage for the LSTM and ResNet solutions, without utilizing any images from the 460 and 640 nm image sets.

Models	(b) 4-Times-Zoom Dataset
ViT-L (P:16 × 16)	72.64%
ViT-L (P:32 × 32)	64.15%
ViT-B (P:16 × 16–C)	70.75%
ViT-B (P:32 × 32–C)	79.24%
Swin-L (P:4 × 4–W:12 × 12)	75.47%
VOLO (D:5)	77.35%
XCIT-L (L:24–P16 × 16)	71.69%
DEIT-L (P16 × 16)	73.58%
MaxViT-L	67.92%

Table 8 provides detailed performance metrics on a per-descriptor-set level for the nine Transformer models for the KNN solution, for multiple distance functions and K values. The table presents the F1-Score for the Euclidian, Cosine, and Bray–Curtis distance functions for a K value of 3, and the F1-Score for 1, 3, 5, and 7 K values for the Euclidian distance function. The results demonstrate that the Euclidian distance function was better by 0.91% and 1.19% than the Bray–Curtis and the Cosine functions, respectively. Additionally, a K value of 3 was 0.18%, 1.13%, and 1.42% more effective than a K value of 1, 5, or 7, respectively. Therefore, these findings validated the choice of a K value of 3 and the Euclidian distance function that were applied in Tables 3–6 for the KNN solution.

Table 8. A performance comparison of the F1-Scores of the different Transformer models for (**a**) Euclidian, Bray–Curtis, Cosine distance functions for the K value of 3; and (**b**) 1, 3, 5, 7 K values for the Euclidian distance function.

Models	(a) Distance Function			(b) K Value			
	Bray–Curtis	Cosine	Euclidian	1	3	5	7
ViT-L (P:16 × 16)	73.96%	72.11%	73.58%	69.81%	73.58%	71.70%	72.64%
ViT-L (P:32 × 32)	71.19%	72.11%	72.64%	76.41%	72.64%	75.47%	73.58%
ViT-B (P:16 × 16–C)	73.96%	71.19%	75.47%	76.42%	75.47%	71.70%	71.70%
ViT-B (P:32 × 32–C)	72.11%	72.11%	77.36%	73.48%	77.36%	73.58%	75.47%
Swin-L (P:4 × 4–W:12 × 12)	76.74%	77.66%	75.47%	76.22%	75.47%	74.53%	75.47%
VOLO (D:5)	75.81%	75.81%	73.58%	75.17%	73.58%	77.36%	74.53%
XCIT-L (L:24–P16 × 16)	74.89%	73.96%	71.70%	73.28%	71.70%	72.64%	72.64%
DEIT-L (P16 × 16)	74.89%	76.74%	77.36%	76.42%	77.36%	74.53%	73.58%
MaxViT-L	73.96%	73.96%	76.42%	75.17%	76.42%	74.53%	74.53%

Ensemble Results

The best classification F1-Score for the KNN, LSTM, and ResNet approaches across all classes were 77.36%, 79.24%, and 73.58%, respectively. The ensemble methodology outlined in Section 2.9 achieved a peak F1-Score across all classes with dataset-b and the LSTM classification solution. The overall F1-Score for this best model combination was 81.13%, while on a per-class level, the F1-Scores were 85.25%, 66.67%, and 78.26% for the "healthy", "botrytis-invisible", and "botrytis-visible" classes, respectively. Table 9 presents all the per-class metrics available for a better insight into the classification performance.

Table 9. Performance metrics on a per-class level from the best ensemble model, based on the LSTM classification approach.

Class	Healthy	Botrytis-Invisible	Botrytis-Visible
Accuracy	83.01%	93.39%	85.84%
Recall	78.78%	87.50%	84.37%
Precision	92.85%	53.84%	72.97%
F1-Score	85.25%	66.67%	78.26%

This ensemble result was obtained by aggregating the predictions of multiple descriptors sets, but not all nine of them. More specifically, the descriptors of four Transformer models proved to yield the best combination for the best outcome. These four models were as follows: the Swin-L (P:4 × 4–W:12 × 12), the ViT-L (P:16 × 16), the VOLO (D:5), and the XCIT-L (L:24–P:16 × 16) models. Also, the aggregation was performed on a per-image-set level. This best result was the combination of only the RGB, CVI, GNDVI, NDVI, and PSRI image sets, while the rest of the image sets, excluding the ones from 460 and 640 nm, were utilized only during training. This proves that the inclusion of the vegetation indices was an important step for improvement. Table 10 shows the combinations of image sets that were tested and their results in the F1-Score metric overall and individually for each class.

For comparison, the best KNN ensemble model yielded an overall F1-Score of 80.18% with the per-class values being 84.90%, 66.67%, and 77.64% for the "healthy", "botrytis-invisible", and "botrytis-visible" classes, respectively. The best ResNet ensemble model delivered an F1-Score of 82.07% overall and 89.28%, 40.00%, and 82.50% on a per-class level. Although the overall F1-Score was a little bit higher for the ResNet compared to the LSTM solution, the low F1-Score (<60%) in the "botrytis-invisible" class was considered a large disadvantage, rendering the ResNet model less viable as a superior alternative.

Table 10. F1-Score (F1) percentages derived from the best ensemble model from the LSTM classification approach on dataset-b for multiple combinations of image set(s) across all classes and on a per-class level. The bold line highlights the image set and the metrics of the best result.

Image Set(s)	Total	Class 0	Class 1	Class 2
RGB	79.25%	85.45%	54.55%	77.50%
CVI	67.92%	76.69%	26.67%	59.38%
GNDVI	70.75%	77.69%	52.63%	63.89%
NDVI	68.87%	75.97%	40.00%	63.49%
NPCI	60.38%	65.93%	14.29%	61.68%
PSRI	69.81%	78.46%	0.00%	66.67%
540	75.47%	85.25%	12.50%	72.97%
775	75.47%	84.75%	14.29%	72.50%
875	75.47%	83.46%	13.33%	74.29%
RGB,540	76.42%	85.22%	13.33%	75.61%
RGB,775,875	76.42%	84.48%	13.33%	76.54%
RGB,540,775,875	76.42%	85.47%	13.33%	75.00%
RGB, GNDVI, NDVI, PSRI	77.36%	85.47%	44.44%	72.73%
RGB, CVI, NDVI, NPCI, PSRI	77.36%	84.75%	52.63%	72.00%
RGB, CVI, GNDVI, NPCI, PSRI	75.47%	83.93%	47.06%	69.88%
RGB, CVI, GNDVI, NDVI, PSRI	81.13%	85.25%	66.67%	78.26%
RGB, CVI, GNDVI, NDVI, NPCI	76.42%	83.19%	50.00%	73.42%
RGB, GNDVI, NDVI, PSRI	74.53%	80.65%	44.44%	71.43%
CVI, GNDVI, NDVI, NPCI, PSRI	77.36%	85.47%	44.44%	72.73%

3.5. Result Visualization

This study aims to tackle the crucial task of plant disease diagnosis in crop management, specifically aiming to precisely detect *B. cinerea* in pepper plants. Figure 11 presents a comparative visualization of the leaf classification, contrasting (a) the manual annotated process and (b) the results produced by the proposed methodology in this study. As shown, the model either successfully detected the pathogen, or, in some cases, it misclassified some "healthy" leaves as infected. It is worth mentioning that these misclassifications could potentially highlight inaccuracies in the manual annotation process and further research could include re-examining and validating some of these cases.

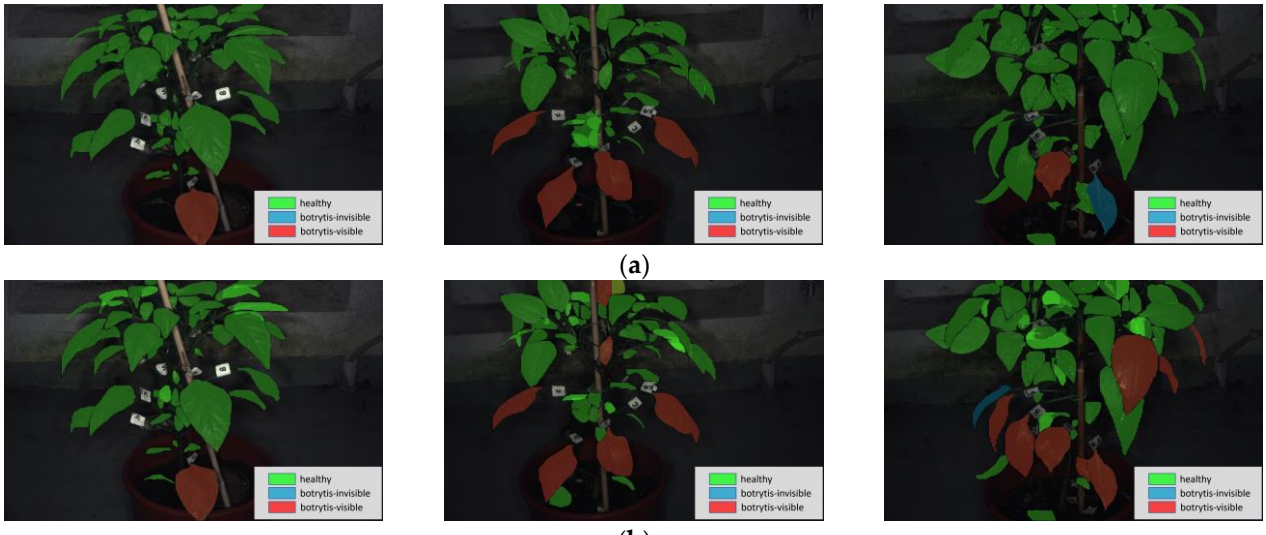

Figure 11. A comparative visualization of three images with their leaf classes drawn on the leaves from (**a**) the manual annotated process and the (**b**) result of the whole methodology of this study. The green leaves represent the "healthy" class, the blue leaves represent the "botrytis-invisible" class, and the red leaves represent the "botrytis-visible" class.

Furthermore, validation was performed by comparing the fungal abundance determined by RT-qPCR with the output results from the proposed approach (Figure 12). Specifically, in Figure 12, images (a–c) demonstrate the precise classification of detached healthy leaves, showing no symptoms of the disease. Images (d–e) highlight the model's capability to detect and classify the presence of *B. cinerea* in the early stages of infection, as early as 1 dpi. Notably, image (f) presents a case observed at 5 dpi, where the model classified a "botrytis-invisible" detached leaf as "botrytis-visible". This classification is in line with the gradual increase in *BcRPL5* expression, confirming the model's sensitivity in detecting the pathogen during the early stages of infection or in latent periods where symptoms have not yet appeared.

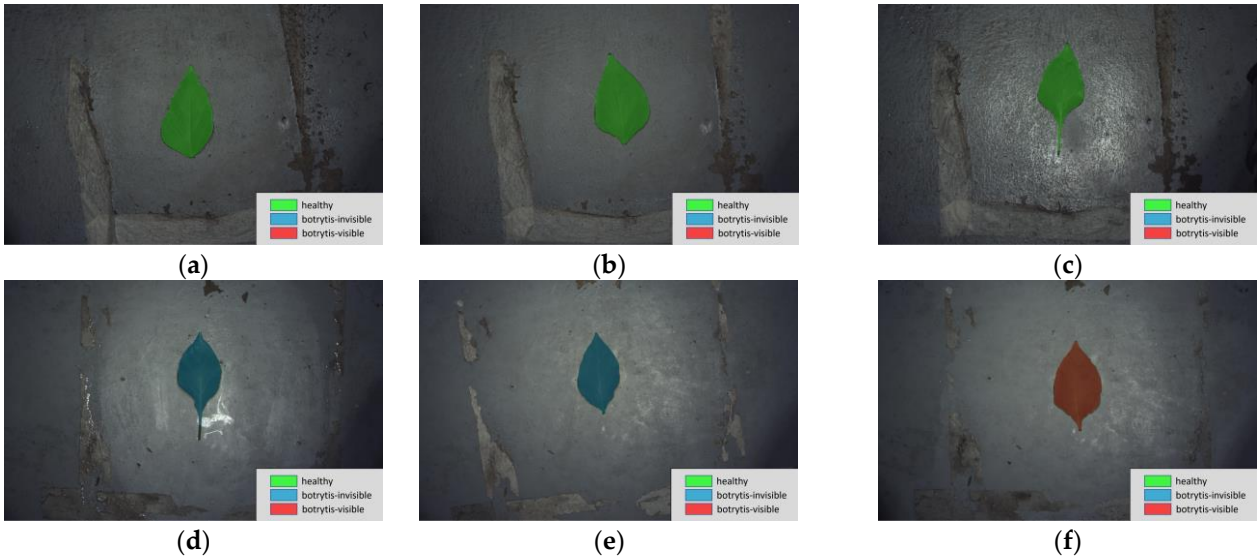

Figure 12. Classification visualization of individual detached leaves used for estimating fungal biomass via RT-qPCR. Images (**a**–**c**) represent mock-inoculated leaves, while images (**d**–**f**) depict *Botrytis cinerea*-inoculated leaves at 1, 2, and 5 dpi, respectively.

4. Conclusions and Future Work

In natural environments, plants encounter a range of biotic and abiotic stressors, including pests, pathogens, and environmental disorders which often produce similar stress symptoms. However, these symptoms alone may not be sufficient to accurately identify their specific causes, making laboratory tests necessary to pinpoint the disease agent—a process that is both time-consuming and labor-intensive. Implementing early-detection strategies for *B. cinerea* in crops is crucial for effective disease management. Additionally, latent infections can lead to sudden outbreaks if not effectively monitored. A promising approach is the use of Artificial Intelligence (AI), particularly DL techniques to enhance detection capabilities and develop protocols for targeted responses, such as the precise application of fungicides or other preventive measures based on AI-generated alerts.

This study focused on detecting *B. cinerea* at various stages of gray mold disease development on pepper plants using 11 image sets (RGB, multispectral, and derived indices) and a two-stage approach: (1) Leaf segmentation using YOLOv11 achieved an mAP50 of 86.4%, effectively locating pepper leaves in images. (2) Leaf classification by stage of infection utilized three zoomed datasets, data augmentation, and Transformer-based descriptors. LSTM-based classification outperformed KNN and ResNet models, with dataset-b yielding the best performance compared to datasets-a and -c. Data augmentation

improved the average F1-Score by 1.07%, while excluding images at 460 and 640 nm improved performance by an average of 2.36%.

The optimal ensemble, which combined descriptors from the Swin-L, ViT-L, VOLO, and XCIT-L models, improved the average F1-Score by 2.38%. Overall, the approach utilized the RGB, CVI, GNDVI, NDVI, and PSRI image sets, and achieved an F1-Score of 81.13%, with class-level performances of 85.25%, 66.67%, and 78.26% for the three classes, respectively. These ensemble results highlight that the solution performed well on the "healthy" class with a high recall, a very high precision, and, consequently, a strong F1-Score. For the "botrytis-invisible" class, the solution showed high recall but average precision, indicating the presence of False Positives. Finally, for the "botrytis-visible" class, the high recall and above-average precision underscored the solution's ability to handle visible symptoms effectively. Overall, the solution demonstrated a solid performance.

The results of this study highlight the importance of monitoring not only visible symptoms, but also the early stages of *B. cinerea* infection, when symptoms are not present, by using molecular markers in combination with DL and multispectral image segmentation. Specifically, the early detection of gray mold infection prior to symptom onset was achieved by quantifying *B. cinerea* biomass in plant tissues, quantified as the relative transcript levels of the *BcRPL5* fungal reference gene, and analyzing the expression of pepper defense-related genes in conjunction with the outputs of the proposed architecture. Notably, the relative transcript levels of *BcRPL5* increased progressively each day during the first 5 dpi, reflecting the pathogen's growth and pathogenicity in the initial infection stages when visible symptoms were not yet present. The contrasting regulation of *DEF1* and *PR1* reflects antagonistic interactions between the SA and JA signaling pathways, highlighting a dynamic plant defense response. The initial downregulation of *DEF1* may benefit the pathogen, while the later activation of *PR1* indicates a strong SA-mediated response to counter the threat. These changes in gene expression, along with variations in *B. cinerea* biomass, illustrate the host's adaptive strategy and the complexity of plant–pathogen interactions during disease progression.

The above integrated approach provides valuable new tools for the identification of gray mold disease in commonly cultivated crops at early and late stages. Notably, this study represents efforts to apply DL techniques to plants developed in controlled environments resembling greenhouse conditions. Moreover, it facilitates timely interventions to protect crop health while reducing the reliance on chemical inputs, as the proposed approach managed to perform the accurate detection of *B.cinerea* on leaves as early as the first dpi.

Computer vision and soft computing techniques, such as those using leaf images, have been utilized by several researchers to automate the detection of plant diseases [50]. According to our study, disease severity progresses more slowly in pepper leaves compared to flowers and fruits, which are more vulnerable to the pathogen. This finding emphasizes the need for the targeted monitoring of these reproductive structures, as they play a crucial role in disease spread and crop losses. Therefore, future studies should focus on investigating the application of these tools to other plant parts beyond leaves. Additionally, further work could focus on areas such as expanding data gathering, manual data validation, and data augmentation. Also, current classification solutions like LSTM and ResNet could be optimized further due to their modular nature, while exploring other classification solutions such as CNN variations and SVMs, or their combinations with the approaches in this study, could enhance detection capabilities.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agriculture15020164/s1, Supplementary Table S1: Primers used in this study.

Author Contributions: Conceptualization, E.M.P.; Data curation, C.K.; Formal analysis, D.K. and E.K.; Funding acquisition, E.M.P.; Investigation, D.K., E.K. and P.C.; Methodology, D.K., E.K., P.C. and E.M.P.; Project administration, E.M.P.; Resources, C.K. and E.M.P.; Software, D.K.; Supervision, E.M.P.; Validation, P.C.; Visualization, D.K. and E.K.; Writing—original draft, D.K. and E.K.; Writing—review and editing, P.C., C.K. and E.M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Data are available on request.

Acknowledgments: This work is partially supported by the Green Deal PestNu project, funded by the European Union's Horizon 2020 research and innovation program under the grant agreement No. 101037128, and partially supported by the E-SPFdigit project, funded by the European Union's Horizon Europe research and innovation program under the grant agreement No. 101157922. We thank E. Fotopoulou and Emilia Markelou from Benaki Phytopathological Institute for their invaluable technical support in molecular experiments. We extend our sincere thanks to Nikolaos Gaikoumoglou for his assistance in capturing the multispectral images and data annotation.

Conflicts of Interest: Author Christos Klaridopoulos was employed by the company iKnowHow S.A. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Molly, F.M.D.; Grant-Downton, R. Botrytis-biology, detection and quantification. In *Botrytis—The Fungus, the Pathogen and Its Management in Agricultural Systems*; Fillinger, S., Elad, Y., Vivier, M., Eds.; Springer: Cham, Switzerland, 2015; pp. 17–34. [CrossRef]
2. Shaw, M.W.; Emmanuel, C.J.; Emilda, D.; Terhem, R.B.; Shafia, A.; Tsamaidi, D.; Emblow, M.; van Kan, J.A.L. Analysis of cryptic, systemic Botrytis infections in symptomless hosts. *Front. Plant Sci.* **2016**, *7*, 625. [CrossRef] [PubMed]
3. Sofianos, G.; Samaras, A.; Karaoglanidis, G. Multiple and multidrug resistance in *Botrytis cinerea*: Molecular mechanisms of MLR/MDR strains in Greece and effects of co-existence of different resistance mechanisms on fungicide sensitivity. *Front. Plant Sci.* **2023**, *14*, 1273193. [CrossRef] [PubMed]
4. Huang, Y.; Duan, C.X.; Lu, M.; Yang, D.F.; Zhu, Z.D. Identification of the pathogens causing chocolate spot on the broad bean. *Plant Prot.* **2012**, *6*, 025.
5. Terentev, A.; Badenko, V.; Shaydayuk, E.; Emelyanov, D.; Eremenko, D.; Klabukov, D.; Fedotov, A.; Dolzhenko, V. Hyperspectral Remote Sensing for Early Detection of Wheat Leaf Rust Caused by *Puccinia triticina*. *Agriculture* **2023**, *13*, 1186. [CrossRef]
6. López, M.M.; Bertolini, E.; Olmos, A.; Caruso, P.; Gorris, M.T.; Llop, P.; Penyalver, R.; Cambra, M. Innovative tools for detection of plant pathogenic viruses and bacteria. *Int. Microbiol.* **2003**, *6*, 233–243. [CrossRef]
7. Schneider, S.; Widmer, F.; Jacot, K.; Koelliker, R.; Enkerli, J. Spatial distribution of Metarhizium clade 1 in agricultural landscapes with arable land and different semi-natural habitats. *Appl. Soil Ecol.* **2012**, *52*, 20–28. [CrossRef]
8. Rezaei, A.; Mahdian, S.; Babaeizad, V.; Hashemi-Petroudi, S.H.; Alavi, S.M. RT-qPCR Analysis of Host Defense-Related Genes in Nonhost Resistance: Wheat-Bgh Interaction. *Russ. J. Genet.* **2019**, *55*, 330–336. [CrossRef]
9. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ. Comput. Inf. Sci.* **2023**, *35*, 757–774. [CrossRef]
10. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is All You Need. *arXiv* **2023**, arXiv:1706.03762. [CrossRef]
11. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929. [CrossRef]
12. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *arXiv* **2021**, arXiv:2103.14030. [CrossRef]
13. Touvron, H.; Cord, M.; Jégou, H. DeiT III: Revenge of the ViT. *arXiv* **2022**, arXiv:2204.07118. [CrossRef]
14. Yuan, L.; Hou, Q.; Jiang, Z.; Feng, J.; Yan, S. VOLO: Vision Outlooker for Visual Recognition. *arXiv* **2021**, arXiv:2106.13112. [CrossRef] [PubMed]

15. El-Nouby, A.; Touvron, H.; Caron, M.; Bojanowski, P.; Douze, M.; Joulin, A.; Laptev, I.; Neverova, N.; Synnaeve, G.; Verbeek, J.; et al. XCiT: Cross-Covariance Image Transformers. *arXiv* **2021**, arXiv:2106.09681. [CrossRef]
16. Tu, Z.; Talebi, H.; Zhang, H.; Yang, F.; Milanfar, P.; Bovik, A.; Li, Y. MaxViT: Multi-Axis Vision Transformer. *arXiv* **2022**, arXiv:2204.01697. [CrossRef]
17. Giakoumoglou, N.; Kalogeropoulou, E.; Klaridopoulos, C.; Pechlivani, E.M.; Christakakis, P.; Markellou, E.; Frangakis, N.; Tzovaras, D. Early detection of *Botrytis cinerea* symptoms using deep learning multi-spectral image segmentation. *Smart Agric. Technol.* **2024**, *8*, 100481. [CrossRef]
18. Christakakis, P.; Giakoumoglou, N.; Kapetas, D.; Tzovaras, D.; Pechlivani, E.-M. Vision Transformers in Optimization of AI-Based Early Detection of *Botrytis cinerea*. *AI* **2024**, *5*, 1301–1323. [CrossRef]
19. Scarboro, C.G.; Ruzsa, S.M.; Doherty, C.J.; Kudenov, M.W. Quantification of gray mold infection in lettuce using a bispectral imaging system under laboratory conditions. *Plant Direct* **2021**, *5*, e00317. [CrossRef]
20. Qasrawi, R.; Amro, M.; Zaghal, R.; Sawafteh, M.; Polo, S.V. Machine Learning Techniques for Tomato Plant Diseases Clustering, Prediction and Classification. In Proceedings of the 2021 International Conference on Promising Electronic Technologies (ICPET), Deir El-Balah, Palestine, 17–18 November 2021; pp. 40–45. [CrossRef]
21. Lorente, Ò.; Riera, I.; Rana, A. Image Classification with Classic and Deep Learning Techniques. *arXiv* **2021**, arXiv:2105.04895. Available online: http://arxiv.org/abs/2105.04895 (accessed on 14 June 2024).
22. Mistry, D.; Banerjee, A. Comparison of Feature Detection and Matching Approaches: SIFT and SURF. *GRD J. Glob. Res. Dev. J. Eng.* **2017**, *2*, 7–13.
23. Mansoori, N.S.; Nejati, M.; Razzaghi, P.; Samavi, S. Bag of visual words approach for image retrieval using color information. In Proceedings of the 2013 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, Iran, 14–16 May 2013; pp. 1–6. [CrossRef]
24. Nagasubramanian, K.; Jones, S.; Singh, A.K.; Sarkar, S.; Singh, A.; Ganapathysubramanian, B. Plant disease identification using explainable 3D deep learning on hyperspectral images. *Plant Methods* **2019**, *15*, 98. [CrossRef] [PubMed]
25. Early Detection of Plant Viral Disease Using Hyperspectral Imaging and Deep Learning. Available online: https://www.mdpi.com/1424-8220/21/3/742 (accessed on 19 September 2024).
26. Kapetas, D.; Kalogeropoulou, E.; Christakakis, P.; Klaridopoulos, C.; Pechlivani, E.M. Multi-spectral image transformer descriptor classification combined with molecular tools for early detection of tomato grey mould. *Smart Agric. Technol.* **2024**, *9*, 100580. [CrossRef]
27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640. [CrossRef]
28. Cunningham, P.; Delany, S.J. k-Nearest Neighbour Classifiers: 2nd Edition (with Python examples). *ACM Comput. Surv.* **2022**, *54*, 1–25. [CrossRef]
29. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385. [CrossRef]
31. Kalogeropoulou, E.; Beris, D.; Tjamos, S.E.; Vloutoglou, I.; Paplomatas, E.J. Arabidopsis β-amylase 3 affects cell wall architecture and resistance against *Fusarium oxysporum*. *Physiol. Mol. Plant Pathol.* **2023**, *124*, 101945. [CrossRef]
32. Kalogeropoulou, E.; Aliferis, K.A.; Tjamos, S.E.; Vloutoglou, I.; Paplomatas, E.J. Combined Transcriptomic and Metabolomic Analysis Reveals Insights into Resistance of Arabidopsis bam3 Mutant against the Phytopathogenic Fungus *Fusarium oxysporum*. *Plants* **2022**, *11*, 3457. [CrossRef]
33. Qcell—Spectral Vision Camera Systems. Available online: https://qcell.tech/ (accessed on 11 September 2024).
34. Petrasch, S.; Silva, C.J.; Mesquida-Pesci, S.D.; Gallegos, K.; van den Abeele, C.; Papin, V.; Fernandez-Acero, F.J.; Knapp, S.J.; Blanco-Ulate, B. Infection Strategies Deployed by *Botrytis cinerea*, *Fusarium acuminatum*, and *Rhizopus stolonifer* as a Function of Tomato Fruit Ripening Stage. *Front. Plant Sci.* **2019**, *10*, 223. [CrossRef]
35. Poveda, J.; Calvo, J.; Barquero, M.; González-Andrés, F. Activation of sweet pepper defense responses by novel and known biocontrol agents of the genus Bacillus against *Botrytis cinerea* and *Verticillium dahliae*. *Eur. J. Plant Pathol.* **2022**, *164*, 507–524. [CrossRef]
36. Wan, H.J.; Yuan, W.; Ruan, M.; Ye, Q.; Wang, R.; Li, Z.; Zhou, G.; Yao, Z.; Zhao, J.; Liu, S.; et al. Identification of reference genes for reverse transcription quantitative real-time PCR normalization in pepper (*Capsicum annuum* L.). *Biochem. Biophys. Res. Commun.* **2011**, *416*, 24–30. [CrossRef] [PubMed]
37. Roboflow: Computer Vision Tools for Developers and Enterprises. Available online: https://roboflow.com/ (accessed on 2 July 2024).
38. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *arXiv* **2020**, arXiv:2001.05566. [CrossRef] [PubMed]
39. Ultralytics. Home. Available online: https://docs.ultralytics.com/ (accessed on 14 June 2024).

40. Zhang, P.; Wu, Y.; Liu, B. Leveraging Local and Global Descriptors in Parallel to Search Correspondences for Visual Localization. *arXiv* **2020**, arXiv:2009.10891. [CrossRef]
41. timm (PyTorch Image Models). Available online: https://huggingface.co/timm (accessed on 3 July 2024).
42. Papers with Code-ImageNet-1K Dataset. Available online: https://paperswithcode.com/dataset/imagenet-1k-1 (accessed on 2 July 2024).
43. Keras: Deep Learning for Humans. Available online: https://keras.io/ (accessed on 2 July 2024).
44. Harikrishnan, N.B. Confusion Matrix, Accuracy, Precision, Recall, F1 Score. Analytics Vidhya. Available online: https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd (accessed on 2 July 2024).
45. Braun, P.; Sutton, J. Infection cycles and population dynamics of *Botrytis cinerea* in strawberry leaves. *Can. J. Plant Pathol.* **1988**, *10*, 133–141. [CrossRef]
46. Williamson, B.; Tudzynski, B.; Tudzynski, P.; Van Kan, J.A.L. *Botrytis cinerea*: The cause of grey mould disease. *Mol. Plant Pathol.* **2007**, *8*, 561–580. [CrossRef]
47. Kim, D.S.; Hwang, B.K. An important role of the pepper phenylalanine ammonia-lyase gene (PAL1) in salicylic acid-dependent signalling of the defence response to microbial pathogens. *J. Exp. Bot.* **2014**, *65*, 2295–2306. [CrossRef]
48. Pieterse, C.M.J.; van der Does, D.; Zamioudis, C.; Leon-Reyes, A.; van Wees, S.C.M. Facilitated Adaptation to Stress: SA-JA Interaction and Crosstalk. *Plant Physiol.* **2012**, *158*, 1743–1754.
49. Thomma, B.P.H.J.; Nurnberger, T.; Joosten, M.H.A.J. Signal Interactions in the Establishment of Plant Immunity. *Curr. Opin. Plant Biol.* **2011**, *14*, 482–489.
50. Vishnoi, V.K.; Kumar, K.; Kumar, B. Plant disease detection using computational intelligence and image processing. *J. Plant Dis. Prot.* **2021**, *128*, 19–53. [CrossRef]