Presymptomatic plant disease detection with PSNet: A low-cost hyperspectral imaging and RGB fusion framework.

Gabriel U. Crabb[1]*, Yiding Zhao[2]*, Nicholas K. Priest†[1], Xi Chen†[2,3], Volkan Cevik†[1]

*These authors contributed equally to this work.

[1] Milner Centre for Evolution, Department of Life Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom

[2] Fudan University, Shanghai, China

[3] Department of Computer Science, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom

† Corresponding authors:

Volkan Cevik (v.cevik@bath.ac.uk)

Xi Chen (xi_chen@fudan.edu.cn)

Nicholas K. Priest (n.priest@bath.ac.uk)

Acknowledgements

This work was funded by the University of Bath

Abstract

Plant pathogens cause major yield losses worldwide, threatening food security and livelihoods. Because early infection is difficult to diagnose, management often relies on prophylactic pesticide use, increasing costs and environmental impact. Here we present PSNet, a multimodal framework that fuses hyperspectral imaging with RGB information for presymptomatic plant disease detection, together with a low-cost, portable hyperspectral camera incorporating a 3D-printed housing and optical mounts, costing under £500. We validate the approach using *Arabidopsis thaliana* infected with the oomycete *Albugo candida*. Imaging at 2 and 4 days post inoculation, prior to visible symptoms, revealed consistent spectral signatures that distinguished infected from healthy plants, while imaging at 6 days post inoculation captured the transition toward early symptom emergence. The most discriminative spectral regions overlapped wavelengths previously associated with plant responses to biotic stress, supporting the biological plausibility of these signatures. On a four-class task (healthy, 2 dpi, 4 dpi, 6 dpi), PSNet achieved 92.7% overall accuracy and 97.1% accuracy for binary healthy versus infected classification. Together, these results demonstrate that presymptomatic detection is feasible under controlled conditions using low-cost hardware and multimodal learning, underscoring the potential of scalable, multimodal systems for early disease monitoring.

Introduction

Plant diseases pose a major and persistent threat to global food security, with pests and pathogens together responsible for up to 40% of annual crop losses worldwide[1,2]. Climate change and expanding global trade are accelerating pathogen spread, altering host-pathogen dynamics and increasing outbreak frequency[3,4]. Despite these intensifying pressures, infection is typically detected only after visible symptoms emerge, when intervention is less effective and yield loss may already be unavoidable. Earlier, presymptomatic detection of infection would enable more targeted management, reduce unnecessary chemical inputs and improve both economic and environmental sustainability.

In practice, disease control often relies on prophylactic or calendar-based applications of fungicides and pesticides. Although these strategies can mitigate outbreak risk, they increase production costs, contribute to environmental contamination and non-target effects, and accelerate the evolution of chemical resistance in pathogen populations[5,6]. Reliable early diagnostics are therefore central to reducing blanket chemical use and enabling precision intervention.

Conventional diagnostic approaches remain constrained. Visual scouting is subjective and detects disease only after symptom development[7]. Molecular assays, including PCR-based tests and related amplification methods, offer high sensitivity but require specialised equipment, controlled laboratory conditions and technical expertise, limiting routine deployment at scale[8,9]. Emerging CRISPR-based diagnostics show promise for rapid detection but remain in early stages of field validation[10]. Collectively, these approaches highlight the tension between diagnostic sensitivity and deployability.

Hyperspectral imaging (HSI) offers a non-destructive alternative capable of detecting subtle physiological changes prior to visible symptoms. By capturing contiguous narrow spectral bands, HSI generates a reflectance spectrum for each pixel, enabling detection of alterations in chlorophyll content, water status and tissue structure associated with biotic stress[11-13]. However, translation into routine use remains limited. Reported detection accuracies vary widely across studies due to differences in sensor design, calibration standards and analytical pipelines[14], and commercial hyperspectral systems remain costly and computationally demanding. As a result, whether accessible, low-cost hyperspectral platforms can reliably support presymptomatic disease detection remains uncertain.

At the analytical level, most hyperspectral classification approaches are unimodal, relying solely on spectral information. Yet infection-related signals are inherently spectral-spatial in nature: biochemical perturbations arise within spatially organised leaf tissues, where spectral variation reflects underlying cellular structure and morphology. Deep learning methods have improved hyperspectral feature extraction[15,16], but systematic integration of hyperspectral and RGB information for presymptomatic disease detection remains underexplored. Whether multimodal fusion can stabilise learning from subtle early-stage spectral signatures is an open question.

To address these challenges, we adapted a compact hyperspectral camera design with a 3D-printed housing and optical mounts[17] for close-leaf presymptomatic imaging. The resulting platform retains a fully 3D-printed structural enclosure, remains repairable, and costs under £500, compared with tens of thousands for commercial systems, while incorporating upgrades that improve acquisition stability and usability. We coupled this accessible hardware with PSNet, a dual-stream deep learning

framework that integrates hyperspectral and RGB information through multimodal fusion.

For biological validation, we employed the *Arabidopsis thaliana-Albugo candida* pathosystem. *Arabidopsis thaliana* provides a genetically tractable model with a well-characterised immune system[18], while *A. candida* exhibits a defined latent phase prior to visible symptom emergence[19]. Plants were imaged at 2 and 4 days post inoculation, corresponding to presymptomatic infection with no visible symptoms, and at 6 days post inoculation to capture the transition toward early symptom emergence, with infection independently verified by trypan blue staining.

In this study, we demonstrate that biologically meaningful infection signatures can be detected during the presymptomatic phase using a low-cost hyperspectral platform and a multimodal learning framework. By integrating accessible hardware with spectral-spatial modelling, we provide a practical approach for early plant disease detection and outline a pathway toward deployable, cost-effective early warning systems for sustainable crop protection.

Materials and Methods

An overview of the experimental workflow used for presymptomatic imaging and analysis is shown in Fig. 1. This workflow integrates controlled plant growth, pathogen inoculation, infection validation, hyperspectral calibration, and co-registered image acquisition.

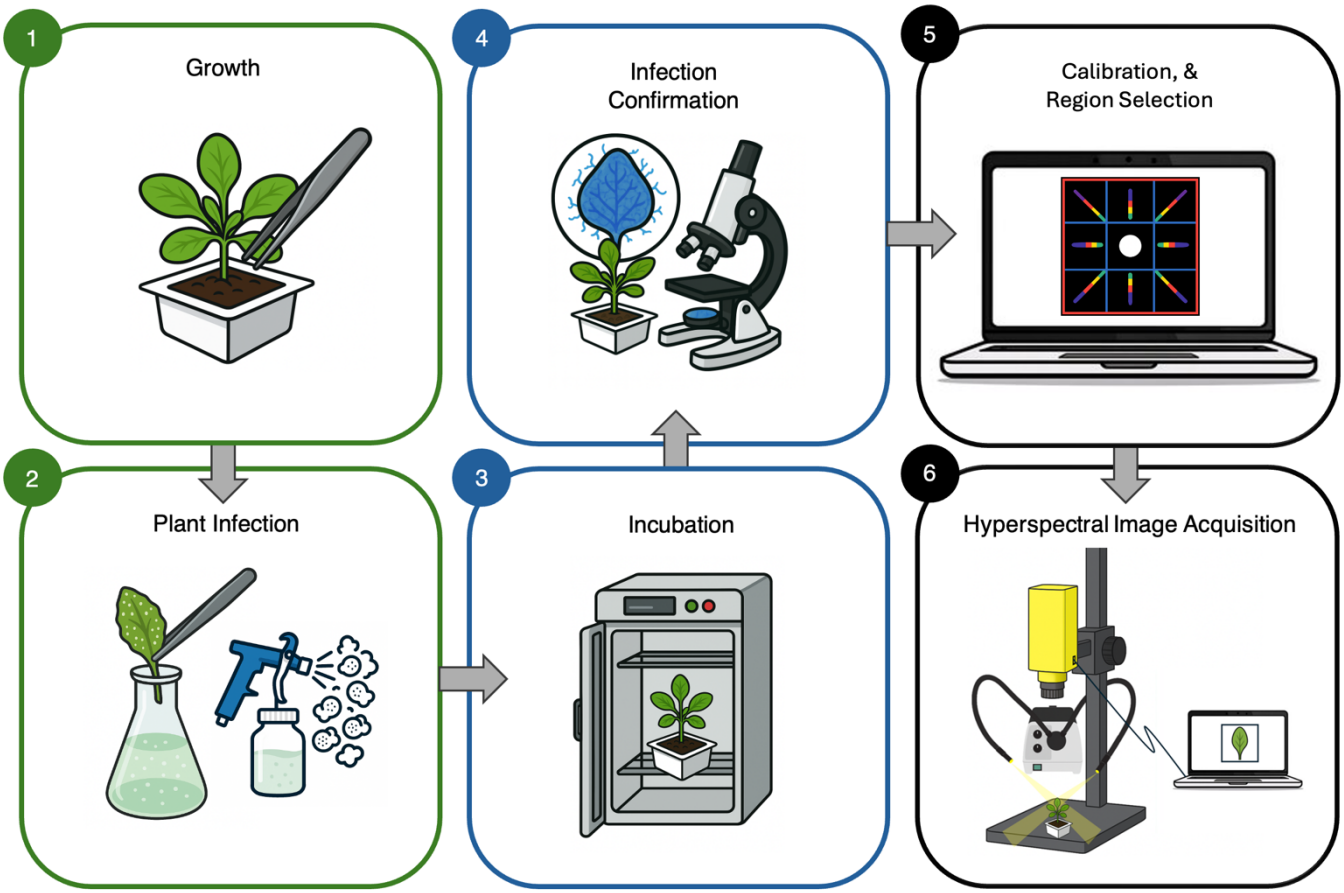

Figure 1. Overview of the data acquisition workflow. *Arabidopsis thaliana* plants were inoculated with *Albugo candida* (controls mock-inoculated), with infection validated by trypan blue staining. Co-registered hyperspectral and RGB images were acquired from leaves 3 and 4 at 2 and 4 days post inoculation (presymptomatic) and at 6 days post inoculation (early symptom transition), using the modified low-cost HSI system under

controlled illumination geometry. Data were wavelength-calibrated, dark-corrected, and processed for downstream modelling.

Plant Material and Growth Conditions

The model organism was fully susceptible *Arabidopsis thaliana* line Col-*eds1.2 wrr7*[20]. This mutant is highly susceptible to the oomycete *Albugo candida* (white blister rust) race 4 isolate AcEM2, enabling reliable infection under controlled conditions and clear contrasts for presymptomatic analysis. Seeds were obtained from a certified stock centre and verified for purity and viability.

Five 6 × 4 cell trays were prepared with a well-drained soil:sand substrate (F2s). An additional pot with the same mix was used for initial germination. Seeds were evenly broadcast and vernalised at 4 °C in darkness for 72 h to promote uniform germination, then transplanted at one plant per cell. Trays stood in water-filled catch trays to maintain uniform moisture by capillarity without waterlogging. Humidity lids were removed at the first true-leaf stage. Unless stated otherwise, conditions were 22 °C**,** 16 h light/8 h dark**, ~**60% relative humidity, and 120-150 µmol m^{-2} s^{-1} photosynthetic photon flux density. Seedlings were used for inoculation approximately three weeks after transplanting, when rosettes bore four to five true leaves.

Pathogen Inoculation and Infection Validation

All spray guns, flasks and bottles were wiped with 100% ethanol before use. Leaves bearing visible *A. candida* pustules were suspended in chilled sterile water; sporangia were dislodged by gentle agitation, filtered through nylon mesh and mixed on ice at a concentration of 10^5 spores ml^{-1}. Controls received sterile water only. For each

run, 96 seedlings were selected (72 inoculated, 24 controls). Both adaxial and abaxial leaf surfaces were misted to run-off. Immediately after spraying, trays were held at 4 °C for 24 h, then placed in a cabinet at 22 °C, 10 h light and 18 °C, 14 h dark until 2 days post-inoculation (dpi), after which standard growth conditions resumed. Leaves were inspected at 7 and 10 dpi for abaxial white pustules. For validation of latent infection, trypan blue staining was performed at 2, 4 and 6 dpi on a reserved subset of leaves at each imaging time point and examined microscopically as described by Furzer et al., 2022.

Hyperspectral Imaging System Design

The hyperspectral imaging (HSI) system was adapted from the open-source 3D-printed design of Salazar-Vazquez and Mendez-Vazquez (2020), retaining a comparable external form factor and overall mass while incorporating updated camera, optical and control components. The original Raspberry Pi Camera Module 2 NoIR was replaced with the Raspberry Pi Camera Module 3 NoIR, integrating a 12-megapixel Sony IMX708 sensor with enhanced sensitivity in the red-edge and near-infrared regions relevant to plant stress and presymptomatic infection detection. The NoIR configuration omits an infrared-blocking filter, preserving spectral response beyond the visible range.

The control platform was upgraded from a Raspberry Pi 3B to a Raspberry Pi 4B (8 GB RAM), improving processing stability and simplifying power delivery via USB-C. A 35 mm C-mount VIS-NIR lens with broadband anti-reflection coating specified for approximately 425-1000 nm was selected to maximise transmission and minimise stray reflections in the red and near-infrared bands. Image acquisition software was migrated

from the legacy raspistill/raspivid utilities to the libcamera stack, enabling stable preview and fine-grained control under current Raspberry Pi OS.

Minor mechanical modifications were introduced to integrate the updated control platform and camera module. Specifically, the outer case was modified with a small USB-C access port to accommodate power delivery to the Raspberry Pi 4B, and the camera base was modified to mount the Raspberry Pi Camera Module 3 NoIR (components 1 and 9 in Figure 2). All other structural elements remained unchanged from the original design. An assembled and exploded view of the final HSI camera system is shown in Figure 2. Complete hardware specifications, mechanical files and wiring diagrams are provided in SI Appendix S1.

A)

B)

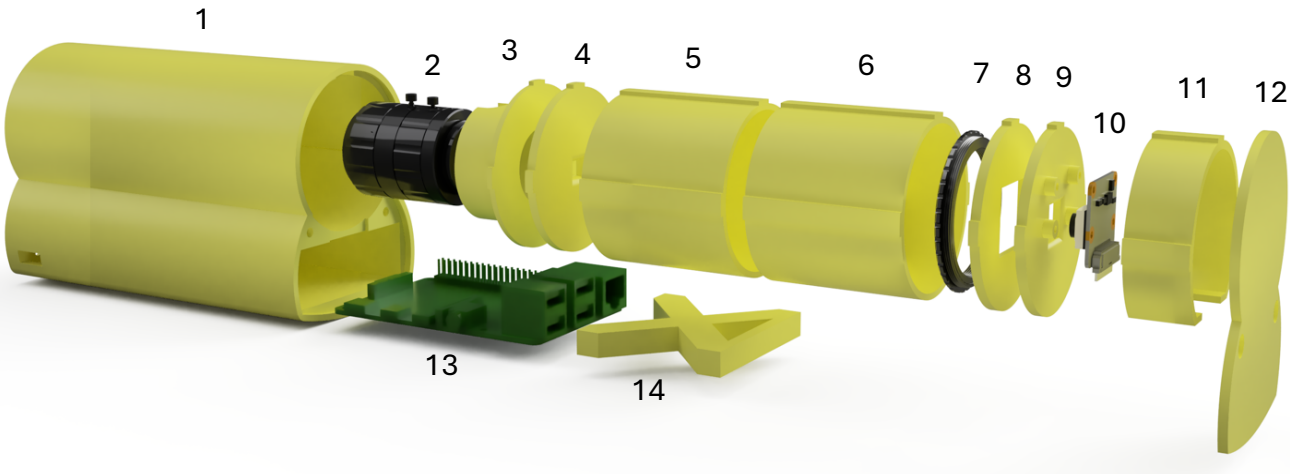

Figure 2 - Assembled view and exploded schematic of the hyperspectral imaging (HSI) camera system. (A) Photograph of the assembled device. (B) Exploded view illustrating the optical, structural, and electronic components and their relative positions. Components are labelled as: **1** case; **2** 35 mm C-mount VIS-NIR fixed focal length lens; **3** front lens; **4** square aperture; **5**-**6** extension tubes; **7** +10 macro 52 mm lens; **8** diffraction grating; **9** camera base; **10** Raspberry Pi Camera Module 3 NoIR (12

MP); **11** seal extension; **12** lid; **13** Raspberry Pi 4 Model B; and **14** X-axis extension. Exploded renders were generated from the CAD models.

Imaging Setup and Data Acquisition

All imaging was conducted in a dark-room to eliminate ambient illumination. Samples were illuminated using a Leica CLS 150 X cold light source (150 W, 3400 K halogen) equipped with dual fibre-optic arms positioned symmetrically at approximately 40° relative to the horizontal sample plane. The illumination geometry was chosen to provide uniform irradiance across the leaf surface while minimising shadowing and specular reflection. The hyperspectral camera was mounted on a copy stand at a fixed working distance of 10 cm above a matte-black background. Illumination sources were allowed to warm for 15 minutes prior to image acquisition to stabilise irradiance. Imaging geometry and illumination parameters were held constant across all sessions.

Spectral Calibration and Preprocessing

Spectral calibration followed the procedure described by Salazar-Vazquez and Mendez-Vazquez (2020). At the beginning of each imaging session, reference measurements were acquired including a fluorescent-lamp spectrum to identify characteristic emission peaks and the zero-order position, an incandescent-lamp spectrum to define the usable spectral range, and a dark frame for sensor noise correction. These references were used to compute a wavelength-to-pixel mapping in squareHSIBetaV1, which was applied uniformly to all hyperspectral datasets. Radiometric parameters, including illumination geometry, exposure time and analogue

gain, were held constant across sessions. Image quality was assessed to confirm correct focus, linear sensor response and absence of saturation. No white reference panel was used; all data were processed as dark-corrected, wavelength-calibrated spectra, consistent with the original Salazar methodology.

Imaging Protocol and Dataset Composition

Image acquisition was performed using the libcamera stack with live preview and autofocus. For each plant and time point, one hyperspectral image cube and one co-registered RGB image were acquired. Exposure time and analogue gain were adjusted to maintain signal levels within the linear response range of the sensor.

A single experimental run was conducted comprising 96 seedlings (72 inoculated and 24 controls). Imaging was performed at 2 and 4 days post inoculation (dpi), corresponding to the fully presymptomatic phase with no visible symptoms, and at 6 dpi to capture the transition toward early symptom emergence. Within each imaging session, plants were imaged in randomised order to minimise temporal drift. Imaging targeted leaves 3 and 4 of each rosette, yielding two co-registered hyperspectral-RGB image pairs per plant per time point. In total, this produced 576 co-registered hyperspectral-RGB image pairs (96 plants × 2 leaves × 3 time points).

Hyperspectral data were transferred to a Linux workstation and assembled into spectral cubes using squareHSIBetaV1. Regions of interest were defined, zero- and first-order diffraction data were extracted, and wavelength calibration and dark-current correction were applied to generate final processed spectral cubes. All acquisitions were automatically logged with plant identifier, treatment and dpi references to ensure full dataset traceability.

Data Pre-Processing

Following wavelength calibration and dark-current correction, all hyperspectral images were standardised on a per-band basis to minimise residual variation arising from illumination and sensor response. Images were cropped and resized to a uniform spatial resolution, and non-leaf background regions were masked. A metadata log recorded plant identifier, treatment, dpi, exposure, gain and calibration file hashes to ensure full dataset traceability.

The final dataset comprised 576 wavelength-calibrated, dark-corrected hyperspectral cubes paired with co-registered RGB images. Dataset partitioning was performed at the leaf-image level. A standalone 10% held-out test set was reserved prior to model development. The remaining 90% of samples were used for five-fold cross-validation, stratified by class to preserve balance within each fold. Reported mean ± standard deviation values reflect performance across the five validation folds. Because splitting was not grouped by plant identifier, leaves 3 and 4 from the same plant could occur in different partitions. Preprocessing parameters were fitted using only the training portion of each fold and were applied unchanged to the corresponding validation and test sets.

Overall Architecture

To address the challenge of detecting subtle pre-symptomatic infection signals, we propose PSNet, a dual-stream deep learning framework designed for synergistic multimodal feature extraction and fusion. As illustrated in Fig. 3, the framework begins with two complementary sensing modalities: high-resolution RGB images and full hyperspectral image (HSI) cubes.

The Spatial Branch processes RGB inputs using a ResNet-based encoder, extracting fine-grained structural, morphological, and textural cues. Even though pre-symptomatic infection is not visually apparent, these high-resolution spatial patterns provide crucial contextual anchors.

The Spectral Branch, depicted in the figure as operating on the HSI cube, performs spectral feature extraction using two complementary representations:3D spectral patches and Average spectral signatures and DWT-transformed spectral curves These representations are fed into the HSNet encoder (shown with CLS token and spectral tokens), which models pixel-level biochemical changes which arise during the earliest infection stages but remain invisible to RGB imaging.

Finally, as shown on the right side of Fig. 3, the RGB feature stream and HSI feature stream are fused to yield a robust multimodal feature representation for classification into Healthy, 2 dpi, 4 dpi, and 6 dpi categories. By integrating high-resolution spatial context with hyperspectral biochemical sensitivity, PSNet learns stable spectral-spatial relations that significantly enhance early infection detection performance.

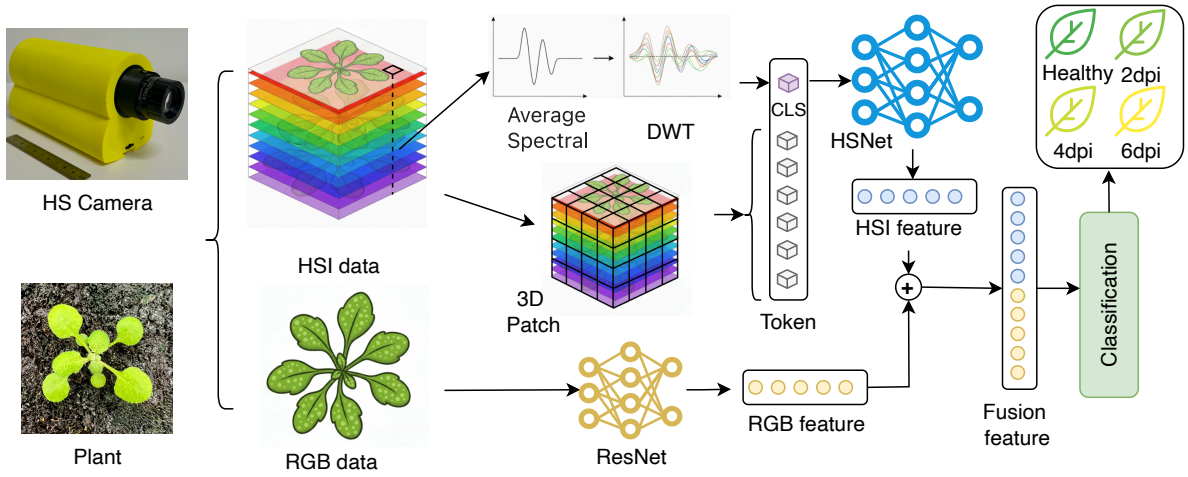

Figure 3. PSNet architecture for multimodal fusion of hyperspectral and RGB information. Co-registered RGB and hyperspectral images from *Arabidopsis thaliana* seedlings are processed by a dual-stream model: the HSNet encoder extracts spectral features from HSI data, and a ResNet backbone extracts spatial features from RGB data. The fused features are classified to determine plant health status, integrating biochemical and morphological cues.

Spatial Branch

The RGB branch adopts a canonical ResNet34 backbone, a widely used architecture in computer vision research that is well suited for hierarchical spatial representation learning. Pre-trained on ImageNet, the network benefits from mature and well-validated convolutional filters that generalize effectively to natural images. Given an input RGB image X_{rgb}, the network processes it through a sequence of convolutional and residual stages, progressively encoding spatial information at increasing levels of abstraction. The residual connections preserve essential information from earlier representations, enabling stable optimization and effective propagation of discriminative spatial cues. As a result, the learned features capture spatial patterns ranging from low-level edges and textures to higher-level morphological structures that are often correlated with visible disease symptoms.

The final feature representation is aggregated via global average pooling (GAP) and passed through a projection head to produce the RGB feature embedding $\boldsymbol{z_{rgb}}$.

Spectral Branch

The spectral branch is designed to model the high dimensionality and strong inter-band correlations inherent to hyperspectral data, while remaining robust under limited sample size. We introduce HSNet, a Transformer-based hyperspectral encoder that represents the hyperspectral cube as a sequence of tokens capturing both global spectral context and local spectral-spatial structure.

Let $\boldsymbol{X}_{hsi} \in \mathbb{R}^{H \times W \times C}$ denote a hyperspectral image cube with spatial dimensions $H \times W$ and C contiguous spectral bands. For each spatial location (i, j), the corresponding spectral reflectance vector is $\boldsymbol{x}_{i,j} \in \mathbb{R}^{C}$.

HSNet constructs its input token sequence from two complementary components computed in parallel from $\boldsymbol{X}_{hsi}$:

1. Wavelet-initialized global spectral token designed to summarise scene-level spectral characteristics, and
2. Set of content tokens derived from non-overlapping 3D spectral-spatial patches that preserve local correlations across neighbouring wavelengths and spatial locations.

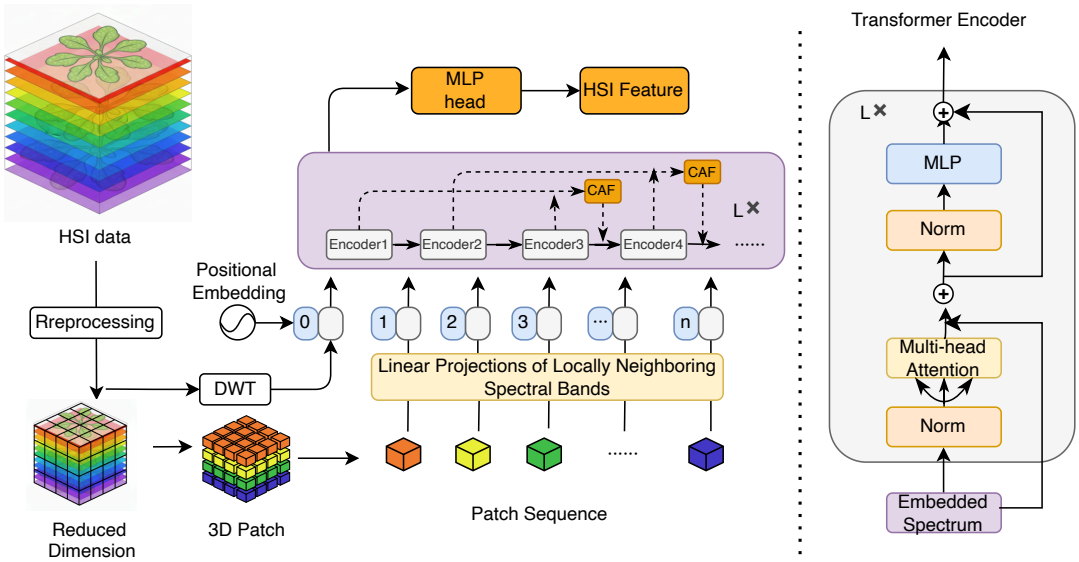

Figure 4. Architecture of the HSNet encoder. The Transformer-based encoder processes hyperspectral image cubes via spectral dimensionality reduction, wavelet transform, non-overlapping 3D patches, and linear projection with positional embedding. Stacked Transformer layers and a cross-layer adaptive fusion mechanism generate final spectral features for downstream analysis.

To provide the model with a strong and physically grounded global prior, we introduce a wavelet-initialized global token $\boldsymbol{x}_{cls}$, designed to encode a compact yet informative representation of the entire hyperspectral scene. In contrast to standard Vision Transformer architectures, where the CLS token is initialized as a learnable random vector with no inherent semantic meaning, our approach constructs this token directly from the spectral properties of the input image. This ensures that the global token carries meaningful, modality-aware information from the outset, facilitating more stable training and enhancing global contextual reasoning.

For each pixel-wise spectral vector $\boldsymbol{x}_{i,j}$, we apply a one-dimensional discrete wavelet transform (DWT) along the spectral dimension. Specifically, an L-level wavelet decomposition is used to separate coarse-scale spectral trends from fine-scale absorption features:

$$\mathcal{W}(\boldsymbol{x}_{i,j}) = \left\{\boldsymbol{a}_{i,j}^{(L)}, \boldsymbol{d}_{i,j}^{(1)}, \dots, \boldsymbol{d}_{i,j}^{(L)}\right\},$$

where $\boldsymbol{a}_{i,j}^{(L)}$ denotes the approximation coefficients at level L, and $\boldsymbol{d}_{i,j}^{(l)}$ are detail coefficients at level l. The wavelet family and decomposition level are fixed across all experiments. In all experiments, a Daubechies-4 (db4) wavelet was used with a fixed decomposition depth of three levels (L = 3). This configuration provides a balance between spectral smoothness and sensitivity to fine-scale absorption features and was held constant across all datasets.

To obtain a spatially invariant global spectral summary, wavelet coefficients are aggregated across all spatial locations using average pooling:

$$\boldsymbol{x}_{\boldsymbol{cls}}^{(\boldsymbol{w})} = \frac{1}{HW}\sum_{u=1}^{H}\sum_{v=1}^{W}\mathcal{W}(\boldsymbol{x}_{i,j})$$

The resulting wavelet-initialized vector summarizes the dominant spectral characteristics of the scene through its coarse-scale components, while retaining subtle biochemical variations encoded in the high-frequency details. Finally, $\boldsymbol{x}_{\boldsymbol{cls}}^{(\boldsymbol{w})}$ is linearly projected into the Transformer latent space and used to initialize the CLS token $\boldsymbol{x}_{cls}$, thereby providing a physically grounded global spectral prior for subsequent cross-modal reasoning.

Having established the global representation through the wavelet-initialized token, we next construct the content tokens that encode local, fine-grained information. We partition the HSI cube $\boldsymbol{X}_{hsi}$ into a sequence of non-overlapping 3D patches. This strategy is superior to simple 2D patching as it explicitly treats local groups of neighboring spectral bands as a cohesive unit, thereby preserving crucial spectral correlation information from the outset.

The resulting data cube is partitioned into a sequence of N non-overlapping 3D patches, $\{\boldsymbol{x}_p^1,\ \boldsymbol{x}_p^2, \ldots, \boldsymbol{x}_p^n\}$, the number of patches along each dimension is:

$$N_c = \lceil\frac{C}{p_c}\rceil, N_h = \left\lceil\frac{H}{p_h}\right\rceil, N_c = \lceil\frac{W}{p_w}\rceil,$$

and the total number of patches is $N = N_c N_h N_w$. This 3D partitioning strategy is a cornerstone of HSNet, as it preserves the local spectral-spatial context within each patch, enabling the downstream Transformer to model both intra-patch fine structures and inter-patch relationships. For a patch index $i = (u, v, w)$, the corresponding 3D patch is obtained by slicing the input tensor:

$$\boldsymbol{X}_p = \boldsymbol{X}_{hsi}[h_p : h_p + P_h, w_p : w_p + P_w, s_p : s_p + P_s].$$

Each patch is then flattened, and passed through a lightweight convolutional embedding module, which is equivalent to a learnable linear operator $\boldsymbol{E}: \mathbb{R}^{p_c \times p_h \times p_w} \rightarrow \mathbb{R}^d$.

This process can be formally expressed as:

$$\boldsymbol{z}_0 = \left[\boldsymbol{x}_{cls}; \boldsymbol{X}_p^1 \boldsymbol{E};\ \boldsymbol{X}_p^2 \boldsymbol{E};\ \ldots;\ \boldsymbol{X}_p^n \boldsymbol{E}\right] + \boldsymbol{E}_{pos}$$

where $\boldsymbol{z}_0 \in \mathbb{R}^{(N+1) \times d}$ denotes the input sequence to the Transformer encoder, consisting of the global token $\boldsymbol{x}_{cls}$ and the sequence of patch embeddings $\boldsymbol{X}_p^i$.

The combined token sequence is processed by a hierarchical Transformer encoder, in which each layer progressively refines the spectral representation via multi-head self-attention and a position-wise feed-forward network. The self-attention mechanism enables the model to capture long-range dependencies across spectral channels, while the hierarchical stacking of layers facilitates the gradual abstraction from local spectral variations to global semantic representations.

To improve information propagation across Transformer layers, we incorporate the Cross-layer Adaptive Fusion (CAF) module proposed by Roy et al. (2019) without modification. CAF introduces mid-range skip connections that enable deeper representations to be adaptively refined using auxiliary information from earlier layers.

Given a shallow-layer feature and a deeper-layer representation, the two vectors are concatenated along the channel dimension and passed through a 1×1 convolution to produce the fused feature $\boldsymbol{z}_l$.

The fused representation$\boldsymbol{z}_l$ is then used as the input to the subsequent Transformer layer. In our framework, the final [CLS] token serves as the global spectral

embedding. While the CAF mechanism itself follows the original design, we find that its integration into a spectral Transformer effectively introduces a form of mid-range memory, which is critical for capturing early biochemical changes in pre-symptomatic leaves.

The final output of the encoder—the [CLS] token from the last layer—serves as the spectral embedding representing the entire hyperspectral patch.

Feature Fusion and Multimodal Classification

The PSNet architecture combines complementary information extracted from RGB and hyperspectral modalities. The RGB branch focuses on spatial and structural cues such as leaf geometry and textural alterations that may precede visible symptoms. In contrast, the hyperspectral branch captures fine-grained spectral variations associated with physiological stress, even when the leaf appears visually healthy. After modality-specific encoding, the spatial feature vector $\boldsymbol{z}_{rgb}$ and the spectral embedding $\boldsymbol{z}_{hsi}$ are concatenated to form a unified representation $\boldsymbol{z}$.

We intentionally adopt concatenation rather than more complex attention-based or bilinear fusion for two reasons. First, concatenation preserves the representational integrity of both modalities without prematurely entangling them—an important property given the subtle and often easily overshadowed spectral cues present in pre-symptomatic stages. Second, more expressive fusion strategies tend to introduce significant parameter overhead and training instability. Concatenation instead provides a stable, high-capacity baseline that enables the classifier to learn cross-modal interactions only when useful.

The fused vector is passed through a lightweight MLP head with GELU activation to produce class logits. Training is performed end-to-end using the cross-entropy objective, which encourages the model to jointly optimize spectral modelling, spatial feature extraction, and multimodal integration.

The network is trained in an end-to-end manner using the cross-entropy loss, which measures the discrepancy between the predicted probability distribution p and the ground-truth label y:

$$\mathcal{L} = -\sum_{k=1}^{K} y_k \log p_k$$

where y_k is a one-hot indicator for class k and p_k denotes the predicted probability for the same class.

Results

Hyperspectral imaging revealed early divergence between infected and healthy *Arabidopsis thaliana* plants. Visual inspection revealed no visible symptoms at 2 and 4 dpi, while at 6 dpi white sporulation became visible on the abaxial leaf surface (Fig. 5). Pathogen infection at all time points was independently confirmed by trypan blue staining. Mean calibrated spectral profiles showed consistent differences beginning at 2 days post inoculation (dpi) and persisting at 4 dpi during the presymptomatic phase, with further divergence at 6 dpi as plants transitioned toward early symptom emergence (Fig. 6). The largest separations occurred in the red-edge region and near-infrared plateau.

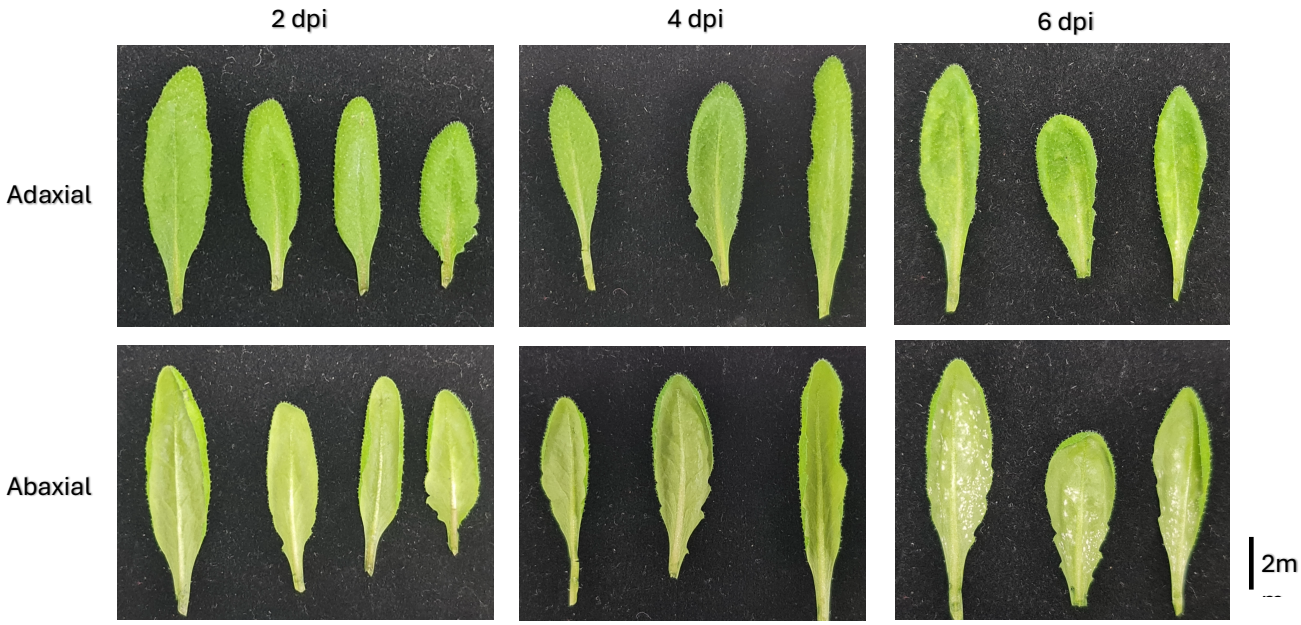

Figure 5. Representative adaxial and abaxial leaf surfaces across infection stages. Representative leaves 3 and 4 of *Arabidopsis thaliana* at 2, 4, and 6 days post inoculation (dpi) with *Albugo candida*. No visible symptoms are apparent at 2 or 4 dpi. White sporulation is visible on the abaxial surface at 6 dpi. Scale bar = 2 mm.

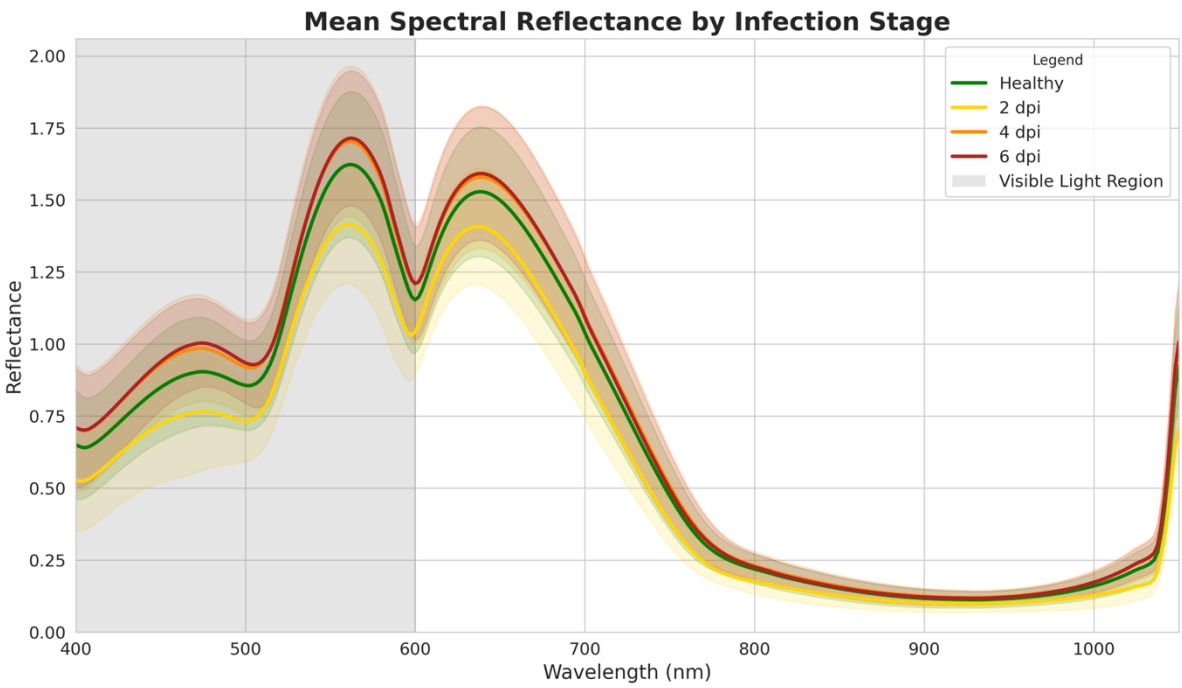

Figure 6. Overlaid mean spectral profiles across infection stages. Mean calibrated spectra are shown for healthy plants (green) and infected plants at 2 dpi (yellow), 4 dpi (orange), and 6 dpi (red). Shaded regions represent variability across samples at each wavelength. The x-axis denotes wavelength (nm) and the y-axis indicates calibrated intensity.

On the four-class classification task spanning healthy plants, two fully presymptomatic stages (2 and 4 dpi), and a transition stage at 6 dpi, the dual-stream PSNet framework achieved an overall accuracy of 92.7 ± 3.9% (Table 1). Presymptomatic stages were reliably detected, with per-class accuracies of 96.9% at 2 dpi and 88.6% at 4 dpi. Binary classification of healthy versus infected plants achieved 97.1% accuracy. The confusion matrix for the standalone 10% held-out test set demonstrated minimal misclassification between adjacent classes (Fig. 7).

Table 1. Performance comparison of baseline models and PSNet using hyperspectral imaging (HSI), RGB imaging, or multimodal fusion. Classical pipelines such as PCA + SVM and PCA + random forest achieved accuracies near chance level for the four-class task (≈55%), whereas deep learning models substantially improved performance. RGB-only ResNet34 achieved 77.5 ± 5.6% accuracy, while HSI-only models performed lower (66.2 ± 8.3%), reflecting the complementary but incomplete information captured by each modality alone. In contrast, PSNet achieved the highest performance across all metrics, including precision, recall, F1-score, and AUROC (0.9918 ± 0.0073).

Model	HSI	RGB	Accuracy	Precision	Recall	F1-score	AUROC (avg.)
PCA + SVM	✓		0.5508 ± 0.0383	0.5373 ± 0.0494	0.5549 ± 0.0389	0.5352 ± 0.0402	0.7914 ± 0.0104
PCA + RF	✓		0.5644 ± 0.0257	0.5677 ± 0.0220	0.5674 ± 0.0266	0.5617 ± 0.0233	0.7905 ± 0.0182
HSNet	✓		0.6145 ± 0.0644	0.5843 ± 0.0740	0.6145 ± 0.0644	0.5818 ± 0.0682	0.8422 ± 0.0272
RGB-ViT		✓	0.3236 ± 0.0848	0.1628 ± 0.1200	0.3236 ± 0.0848	0.1946 ± 0.1118	0.5933 ± 0.1355
HSI-ViT	✓		0.5455 ± 0.0381	0.4046 ± 0.0655	0.5455 ± 0.0381	0.4399 ± 0.0595	0.7845 ± 0.0149
RGB-ResNet		✓	0.7754 ± 0.0561	0.7875 ± 0.0571	0.7754 ± 0.0561	0.7672 ± 0.0662	0.8800 ± 0.0341
HSI- ResNet	✓		0.7164 ± 0.0317	0.7272 ± 0.0289	0.7164 ± 0.0317	0.6950 ± 0.0379	0.9021 ± 0.0199
Both-ResNet	✓	✓	0.7709 ± 0.0751	0.7758 ± 0.0762	0.7709 ± 0.0751	0.7636 ± 0.0729	0.9039 ± 0.0404
Both-Vit	✓	✓	0.5345 ± 0.0810	0.4926 ± 0.1194	0.5345 ± 0.0810	0.4837 ± 0.0959	0.7720 ± 0.0435
PSNet (ours)	✓	✓	**0.9273 ± 0.0398**	**0.9317 ± 0.0384**	**0.9273 ± 0.0398**	**0.9273 ± 0.0395**	**0.9918 ± 0.0073**

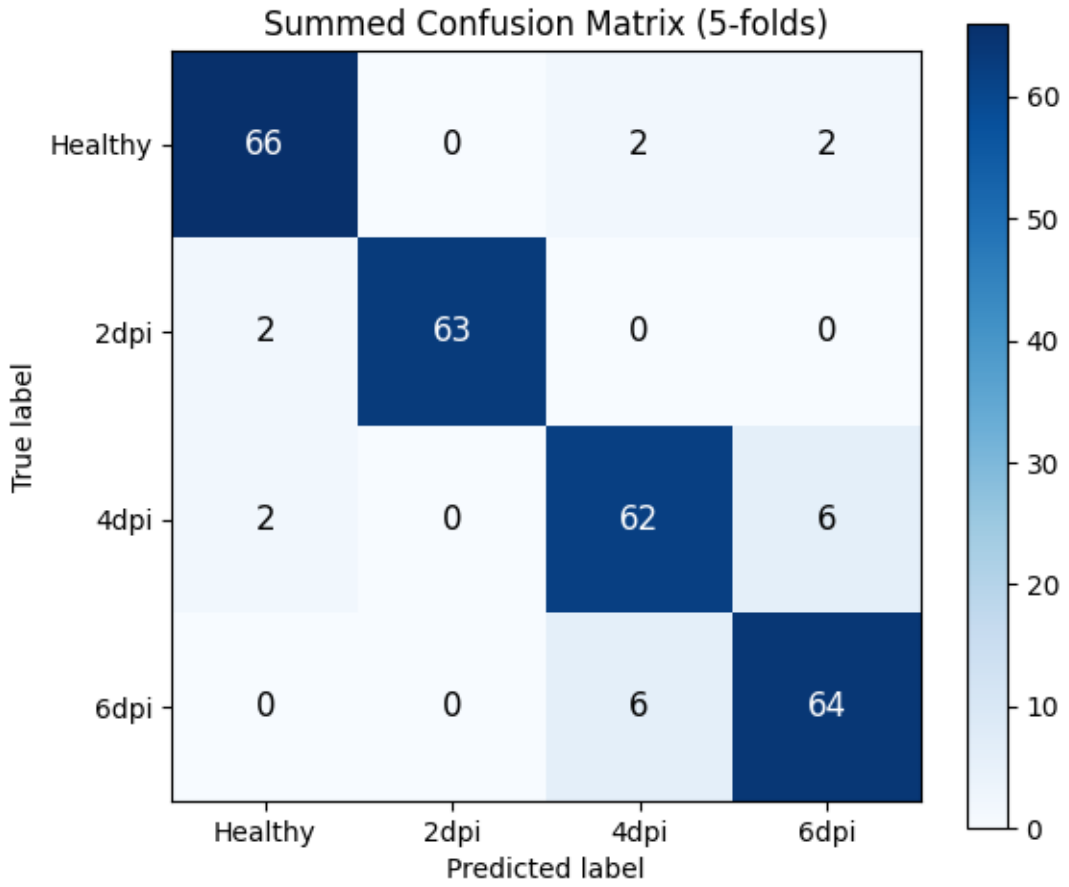

Figure 7. Confusion matrix for the four-class classification task on the held-out test set. Values represent aggregated predictions for Healthy, 2 dpi, 4 dpi, and 6 dpi classes on the standalone 10% test partition.

Beyond overall benchmarking, we examined whether the observed gains could be attributed primarily to ImageNet pretraining rather than multimodal representation learning. To this end, we repeated training with randomly initialised encoders while keeping all other settings unchanged (Table 2).

Table 2. Influence of encoder pretraining. Benchmarking against baseline approaches confirmed that PSNet consistently outperformed unimodal and classical methods (Table 1). Ablation studies further demonstrated that key architectural components of the HSNet encoder were critical for optimal performance (Table 3). Removal of either the wavelet-initialised global token or the cross-layer adaptive fusion (CAF) mechanism led to substantial reductions in classification accuracy, confirming their contribution to stable and discriminative spectral representation learning.

Model	HSI	RGB	Accuracy	Precision	Recall	F1-score	AUROC (avg.)
ResNet(NoPretrain)		✓	0.6545 ± 0.0527	0.6473 ± 0.1108	0.6545 ± 0.0527	0.6303 ± 0.0754	0.8837 ± 0.0172
ResNet(Pretrain)		✓	0.7754 ± 0.0561	0.7875 ± 0.0571	0.7754 ± 0.0561	0.7672 ± 0.0662	0.8800 ± 0.0341
PS**Net**(NoPretrain)	✓	✓	0.8255 ± 0.0337	0.8353 ± 0.0406	0.8255 ± 0.0337	0.8255 ± 0.0366	0.9388 ± 0.0244
PSNet (Pretrain)	✓	✓	**0.9273 ± 0.0398**	**0.9317 ± 0.0384**	**0.9273 ± 0.0398**	**0.9273 ± 0.0395**	**0.9918 ± 0.0073**

Table 3. Results of the ablation study. The full PSNet model outperformed all reduced variants, with the largest performance drop observed when the wavelet module was removed, followed by removal of CAF and the 3D patch embedding strategy. Across all comparisons, multimodal fusion consistently improved performance relative to single-modality models, reducing confusion between adjacent presymptomatic stages.

Model Variant	Accuracy	Precision	Recall	F1-Score	AUROC (avg.)
PSNet	**0.9273 ± 0.0398**	**0.9317 ± 0.0384**	**0.9273 ± 0.0398**	**0.9273 ± 0.0395**	**0.9918 ± 0.0073**
w/o CAF	0.9055 ± 0.0267	0.9119 ± 0.0237	0.9055 ± 0.0267	0.9054 ± 0.0261	0.9832 ± 0.0113
w/o Wavelet Module	0.8727 ± 0.0501	0.8818 ± 0.0481	0.8727 ± 0.0501	0.8727 ± 0.0500	0.9780 ± 0.0090
w/o 3D Patch	0.8509 ± 0.0506	0.8642 ± 0.0420	0.8509 ± 0.0506	0.8499 ± 0.0506	0.9717 ± 0.0190

Discussion

This study demonstrates that presymptomatic infection signatures (2 and 4 dpi) can be captured in hyperspectral data and reliably classified when fused with RGB information, with 6 dpi included as a transition stage toward early symptom emergence. Across two fully presymptomatic time points at 2 and 4 days post inoculation and a transition time point at 6 days post inoculation, mean calibrated reflectance profiles diverged consistently between infected and healthy plants, and the PSNet framework separated infection stages with high accuracy. These time points were selected to span fully presymptomatic infection (2 and 4 dpi) and the transition toward early symptom emergence (6 dpi), aligning with the presymptomatic sensitivity window reported by hyperspectral studies in other host-pathogen systems[21,22]. Together, these findings support hyperspectral imaging as a practical route to early plant disease detection within an appropriate analytical and acquisition framework.

The spectral changes observed in the red edge and near infrared regions align with well-established stress sensitive wavelengths[11-13]. The red edge is strongly influenced by chlorophyll concentration and photosynthetic apparatus status, whereas the near infrared plateau is shaped by internal leaf structure and scattering properties[14]. The progressive amplification of spectral differences from 2 to 6 dpi is consistent with gradual shifts in pigment composition and tissue microstructure prior to symptom expression. Although the present analyses were not designed to assign causal physiological mechanisms, the observed patterns are biologically plausible and provide credible evidence for an early infection signature that precedes visible disease.

On a four-class task spanning healthy and presymptomatic stages, PSNet achieved an overall accuracy of 92.7 ± 3.9 percent, outperforming all unimodal and

classical baselines. For the binary task of distinguishing healthy from infected plants, accuracy increased further to 97.1 percent, underscoring the sensitivity of the fused representation to early physiological perturbations. Compared with traditional pipelines such as PCA combined with SVM or random forest classifiers, which rely on handcrafted dimensionality reduction and shallow decision boundaries, PSNet exhibited a performance margin exceeding 36 percent. This substantial gap highlights the limitations of linear feature projections for capturing the nonlinear spectral spatial dependencies that characterise early plant disease.

An unexpected but informative result is that RGB-only models outperformed HSI-only models on this dataset, despite hyperspectral data theoretically subsuming RGB information. This observation does not contradict the added value of hyperspectral sensing but rather reflects practical constraints of learning from high-dimensional spectral data under limited sample size.

RGB models benefit from strong spatial inductive biases and large-scale pretraining on natural image datasets, enabling stable extraction of geometric and textural features even when disease symptoms are not visually apparent. These features may encode indirect correlates of infection stage, such as subtle growth-related or morphological variations, without being specific biomarkers of disease.

In contrast, hyperspectral models must learn meaningful spectral representations from scratch in a high-dimensional space, which increases statistical complexity and sensitivity to noise. Under these conditions, spectral information alone may be insufficiently constrained to outperform pretrained RGB encoders. Importantly, the complementary nature of the two modalities is evidenced by the substantial

performance gains achieved through multimodal fusion, indicating that hyperspectral features contribute disease-relevant information not captured by RGB alone.

The substantial performance gain observed with multimodal fusion arises from the complementary strengths of the two modalities rather than simple information aggregation. RGB features constrain the hypothesis space by enforcing spatial and structural regularity, while hyperspectral features contribute disease-specific biochemical sensitivity within this constrained space. This interaction leads to a multiplicative improvement in representational stability and discriminative power, explaining the disproportionately large increase in accuracy achieved by PSNet relative to either modality alone.

Ablation analysis further clarified the contributions of individual architectural components. The wavelet initialised global token improved performance by introducing a denoised spectral prior that summarised global biochemical trends while reducing redundant band level noise. The cross layer adaptive fusion mechanism stabilised hierarchical feature formation by aligning representations across encoder depths, preserving fine grained spectral information while integrating higher level abstractions. Removal of either component resulted in substantial performance degradation, demonstrating that PSNet's gains arise from principled representational design rather than parameterisation alone.

The modality comparison provides useful guidance for future system design. While RGB sensing offers low-cost spatial context and benefits from pretrained feature extractors, hyperspectral data contribute complementary biochemical information that is critical for early stress detection, as shown previously in crop monitoring studies[12,23]. Their fusion consistently reduced confusion between adjacent presymptomatic stages,

supporting a practical design principle for field systems: combine inexpensive RGB imaging for spatial anchoring with targeted spectral measurements to capture subtle physiological changes.

Beyond algorithmic performance, the compact and portable design of the hyperspectral platform supports potential translational use. The system retains near-identical dimensions and weight to the Salazar-Vazquez and Mendez-Vazquez[17] build, allowing for handheld or mobile operation in controlled settings. USB-C power compatibility enables operation from laptops or portable battery packs, and the modular 3D-printed housing facilitates on-site repair and low-cost replacement of damaged components. Together with the observed robustness of spectral signatures to moderate illumination variation in this dataset, these features indicate that the platform is well suited for deployment in greenhouse and other controlled or semi-controlled environments. Extension to fully open field conditions would require additional consideration of environmental variability and the use of appropriate calibration standards, as noted previously[24,25].

Several limitations should be acknowledged. First, this study focused on a single host pathogen system under controlled conditions. Generalisation to diverse crops, developmental stages and environmental contexts remains to be tested. Second, the dataset did not include confounding abiotic stresses such as drought or nutrient deficiency, which are known to induce overlapping spectral responses[7,26]. Third, only one optical configuration and illumination geometry were evaluated, and robustness to natural lighting, leaf orientation and motion remains an open challenge[14]. Finally, infection progression was assessed at discrete time points rather than continuously, limiting temporal resolution.

These limitations point to clear directions for future work. Multi location datasets spanning genotypes, growth stages and management regimes will be required to assess transfer performance under realistic conditions. Self-supervised or weakly supervised pretraining on unlabelled hyperspectral data could reduce sample complexity, analogous to ImageNet pretraining for RGB models[27]. Model interpretability should be strengthened through spectral attribution maps, class prototypes and physiologically grounded indicators such as red edge shifts, enabling alignment between model predictions and known plant processes[12,27,28]. Additional modalities, including thermal or fluorescence imaging, may further enhance robustness in uncontrolled environments.

In summary, presymptomatic spectral signatures of infection were detectable at 2 and 4 days post inoculation, several days before visible symptoms. Inclusion of an additional 6 dpi transition time point enabled robust staging as plants approached early symptom emergence, with high classification accuracy achieved using a dual-stream framework integrating hyperspectral and RGB information. The biological plausibility of the affected spectral regions, the substantial benchmarking gains over unimodal approaches, and the practical cost and portability of the imaging platform together indicate a realistic pathway toward accessible early warning tools for plant disease monitoring. Future efforts should prioritise broader validation, and interpretable multimodal models to translate these findings into routine agricultural practice.

Author Contributions

The study was initiated by V.C. as part of G.U.C.'s doctoral research. The hyperspectral imaging system and experimental workflow were designed and executed by G.U.C., including data acquisition, spectral calibration, preprocessing, and dataset construction, under the supervision of N.K.P. and V.C. G.U.C. performed exploratory modelling and preliminary computational analyses, drafted the manuscript, and led the biological interpretation in collaboration with co-authors. Y.Z. developed the multimodal machine learning architecture and implemented PSNet, including the training and validation framework. X.C. provided supervision and methodological guidance on the machine learning and computational framework. All authors contributed to discussion, interpretation, and manuscript revision.

References

1 Savary, S. *et al*. The global burden of pathogens and pests on major food crops. *Nature Ecology & Evolution* **3**, 430-439 (2019). https://doi.org:10.1038/s41559-018-0793-y

2 FAO. The impact of disasters and crises on agriculture and food security: 2021. (Rome, 2021).

3 Hossain, M. M. *et al*. Plant disease dynamics in a changing climate: impacts, molecular mechanisms, and climate-informed strategies for sustainable management. *Discover Agriculture* **2**, 132 (2024). https://doi.org:10.1007/s44279-024-00144-w

4 Ristaino, J. B. *et al*. The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences* **118**, e2022239118 (2021). https://doi.org:doi:10.1073/pnas.2022239118

5 Fisher, M. C., Hawkins, N. J., Sanglard, D. & Gurr, S. J. Worldwide emergence of resistance to antifungal drugs challenges human health and food security. *Science* **360**, 739-742 (2018). https://doi.org:10.1126/science.aap7999

6 Lucas, J. A., Hawkins, N. J. & Fraaije, B. A. The evolution of fungicide resistance. *Advances in Applied Microbiology*

90, 29-92 (2015). https://doi.org:10.1016/bs.aambs.2014.09.001

7 Mahlein, A. K. Plant Disease Detection by Imaging Sensors - Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Dis* **100**, 241-251 (2016). https://doi.org:10.1094/pdis-03-15-0340-fe

8 Lau, H. Y. & Botella, J. R. Advanced DNA-Based Point-of-Care Diagnostic Methods for Plant Diseases Detection. *Frontiers in Plant Science* **Volume 8 - 2017** (2017). https://doi.org:10.3389/fpls.2017.02016

9 Notomi, T., Mori, Y., Tomita, N. & Kanda, H. Loop-mediated isothermal amplification (LAMP): principle, features, and future prospects. *J Microbiol* **53**, 1-5 (2015). https://doi.org:10.1007/s12275-015-4656-9

10 Hak, H. *et al.* Rapid on-site detection of crop RNA viruses using CRISPR/Cas13a. *Journal of Experimental Botany* (2024). https://doi.org:10.1093/jxb/erae495

11 Lowe, A., Harrison, N. & French, A. P. Hyperspectral image analysis techniques for the detection and classification of the early onset of plant disease and stress. *Plant Methods* **13**, 80 (2017). https://doi.org:10.1186/s13007-017-0233-z

12 Mahlein, A. K., Kuska, M. T., Behmann, J., Polder, G. & Walter, A. Hyperspectral Sensors and Imaging Technologies in Phytopathology: State of the Art. *Annu Rev Phytopathol* **56**, 535-558 (2018). https://doi.org:10.1146/annurev-phyto-080417-050100

13 Terentev, A., Dolzhenko, V., Fedotov, A. & Eremenko, D. Current State of Hyperspectral Remote Sensing for Early Plant Disease Detection: A Review. *Sensors* **22**, 757 (2022).

14 Paulus, S. & Mahlein, A.-K. Technical workflows for hyperspectral plant image assessment and processing on the greenhouse and laboratory scale. *GigaScience* **9** (2020). https://doi.org:10.1093/gigascience/giaa090

15 Chen, Y.-N., Thaipisutikul, T., Han, C.-C., Liu, T.-J. & Fan, K.-C. Feature Line Embedding Based on Support Vector Machine for Hyperspectral Image Classification. *Remote Sensing* **13**, 130 (2021).

16 Roy, S., Krishna, G., Dubey, S. R. & Chaudhuri, B. HybridSN: Exploring 3-D-2-D CNN Feature Hierarchy for Hyperspectral Image Classification. *IEEE Geoscience and Remote Sensing Letters* **17**, 277-281 (2019). https://doi.org:10.1109/LGRS.2019.2918719

17 Salazar-Vazquez, J., Mendez-Vazquez, A. A plug-and-play Hyperspectral Imaging Sensor using low-cost equipment. *HardwareX* **7** (2020). https://doi.org:https://doi.org/10.1016/j.ohx.2019.e00087

18 Koornneef, M. & Meinke, D. The development of Arabidopsis as a model plant. *Plant J* **61**, 909-921 (2010). https://doi.org:10.1111/j.1365-313X.2009.04086.x

19 Borhan, M. H. *et al.* WRR4, a broad-spectrum TIR-NB-LRR gene from Arabidopsis thaliana that confers white rust resistance in transgenic oilseed Brassica crops.

Mol Plant Pathol **11**, 283-291 (2010). https://doi.org:10.1111/j.1364-3703.2009.00599.x

20 Parkes, T. *From Recognition to Susceptibility: Functional characterization of Plant- specific LIM-domain containing proteins in plant-microbe interactions* PhD thesis, University of Bath, (2020).

21 Zhu, H. *et al.* Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-learning Classifiers. *Scientific Reports* **7**, 4125 (2017). https://doi.org:10.1038/s41598-017-04501-2

22 Zhang, X., Vinatzer, B. A. & Li, S. Hyperspectral imaging analysis for early detection of tomato bacterial leaf spot disease. *Scientific Reports* **14**, 27666 (2024). https://doi.org:10.1038/s41598-024-78650-6

23 Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K. & Sun, X. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture* **222**, 109037 (2024). https://doi.org:https://doi.org/10.1016/j.compag.2024.109037

24 Sousa, J. J. *et al.* UAV-Based Hyperspectral Monitoring Using Push-Broom and Snapshot Sensors: A Multisite Assessment for Precision Viticulture Applications. *Sensors* **22**, 6574 (2022).

25 Detring, J., Barreto, A., Mahlein, A.-K. & Paulus, S. Quality assurance of hyperspectral imaging systems for neural network supported plant phenotyping. *Plant Methods* **20**, 189 (2024). https://doi.org:10.1186/s13007-024-01315-y

26 Singh, B. K. *et al.* Climate change impacts on plant pathogens, food security and paths forward. *Nature Reviews Microbiology* **21**, 640-656 (2023). https://doi.org:10.1038/s41579-023-00900-7

27 Hong, D. *et al.* SpectralFormer: Rethinking Hyperspectral Image Classification With Transformers. *IEEE Transactions on Geoscience and Remote Sensing* **PP**, 1-1 (2021). https://doi.org:10.1109/TGRS.2021.3130716

28 Wang, N. *et al.* Key Challenges in Plant Pathology in the Next Decade. *Phytopathology* **114**, 837-842 (2024). https://doi.org:10.1094/phyto-04-24-0137-kc