

PCR-Validated Comparative Deep Learning Analysis of CNN and YOLO Models for UAV-Based multi-spectral Detection of Sugarcane White Leaf Disease (WLD)

Dahami Samindi Wickramasinghe

dahami.s@ruh.ac.lk

University of Ruhuna

Research Article

Keywords: convolutional neural networks (CNN), deep learning, multi-spectral imagery, precision agriculture, sugarcane white leaf disease, UAV remote sensing

Posted Date: September 22nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11111373/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

PCR-Validated Comparative Deep Learning Analysis of CNN and YOLO Models for UAV-Based multi-spectral Detection of Sugarcane White Leaf Disease (WLD)

Dahami Samindi Wickramasinghe^{1*}, W B M A C. Bandara², Guttula Yugantha Jayasinghe², Thilanka Ariyawansa³, Sumedha Thushari⁴

*¹Department of Crop Science, Faculty of Agriculture,
University of Ruhuna, Sri Lanka*

*²Department of Agricultural Engineering and Environmental Technology, Faculty of Agriculture,
University of Ruhuna, Sri Lanka*

³Department of Agricultural Technology, Faculty of Technology, University of Colombo, Sri Lanka

⁴Division of Crop Protection, Sugarcane Research Institute, Udawalawe, Sri Lanka

Corresponding author: Dahami Samindi Wickramasinghe, [Department of Crop Science, Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka]; e-mail: [dahami.s@ruh.ac.lk]

ORCID: [0009-0006-9173-9178]

Acknowledgments: The technical support provided by the Sugarcane Research Institute, Sri Lanka, in performing the PCR-based molecular diagnosis of the sugarcane samples used in this study is gratefully acknowledged.

Key points

- A dataset of 115 sugarcane plants (102 healthy and 13 white leaf disease) was validated with PCR to evaluate CNN, YOLOv5s, and YOLOv8s for the UAV multi-spectral detection of white leaf disease.
- The CNN achieved 98.85% precision, 95.56% recall, and 97.17% F1-score, performing competitively against both YOLO variants and previously published WLD detection results to date under the conditions of this study.
- For YOLO-based detection, the green waveband (560 nm) proved to be the most discriminative single band, whereas the CNN made efficient use of all bands except the blue (450 nm).
- All model predictions were evaluated with PCR-confirmed disease ground truth, which is rare in studies using an UAV for disease detection.
- The findings are preliminary, and apply to a single season and limited dataset; they need external validation to larger, multi-season, multi-cultivar datasets before they are deployed operationally.

Impact: The study has proven that a combination of UAV-based multi-spectral imaging and deep learning is technically feasible for detecting sugarcane white leaf disease in early stages, and can assist with spatial disease monitoring as part of precision agriculture with technologies.

Keywords: convolutional neural networks (CNN); deep learning; multi-spectral imagery; precision agriculture; sugarcane white leaf disease; UAV remote sensing

Abstract

Context: Sugarcane White Leaf Disease (WLD), caused by phytoplasma of the 16SrXI group, threatens global sugarcane production and can cause total yield loss in susceptible varieties. For effective disease management, early and scalable detection is crucial.

Aims: This study presents a PCR-validated comparative analysis of three deep learning models, YOLOv5s, YOLOv8s, and a TensorFlow/Keras-based Convolutional Neural Network (CNN), for automated WLD detection using UAV-acquired multi-spectral imagery.

Methods: The plant samples were collected from 13 sugarcane fields at three locations in Sri Lanka, comprising a total of 115 samples, including 102 healthy and 13 diseased, confirmed by PCR. Regions of interest were extracted from orthomosaics across six imaging channels (five multi-spectral bands plus one broadband RGB channel). Class imbalance was addressed by creating a balanced dataset of 1,800 images, which were split into training, validation, and testing sets using stratified random sampling; augmentation was performed before partitioning, which left a residual risk of data leakage as mentioned in the limitations, with the small number of diseased plants. Models were evaluated per spectral band and on the full multi-spectral dataset using precision, recall, F1-score, and mean Average Precision (mAP@50 and mAP@50–95; YOLO models only).

Key Results: The CNN achieved 98.85% precision, 95.56% recall, and 97.17% F1-score, performing competitively against previously published WLD detection results to date. The green waveband (560 nm) proved to be most discriminative for YOLO models and the CNN used all wave bands apart from the blue band (450 nm). Ground truth was confirmed by PCR and all predictions were tested against this.

Conclusion: The results indicate that automated WLD detection using UAV multi-spectral imaging and deep learning, specifically CNN models, is technically feasible for precision agriculture, as it demonstrates satisfactory performance in this specific context with limited number of datasets and a single season, and emphasizes the need for optimization of the spectral bands.

Implications and Impacts: External validation is needed on larger/longer season/more varieties before operational deployment is recommended.

Statement and Declarations

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Funding: The authors did not receive any financial support from any organization for the submitted work.

Ethics Approval: Not applicable. This study did not involve human participants or animals; leaf tissue was sampled non-destructively from commercial sugarcane plants with permission from the respective plantation management (Gal-Oya, Pelwatte, and Ethimale Plantations in Sri Lanka).

Plant Ethics: The sampled sugarcane plants (varieties Co 775 and SL 96383) are commercially cultivated, non-endangered, non-genetically-engineered cultivars; no new plant taxa or nomenclatural acts are reported, no protected or endangered species were collected, and the Nagoya Protocol and voucher-specimen deposition requirements are therefore not applicable to this study. Sugarcane White Leaf Disease phytoplasma is endemic and widely documented in Sri Lanka (Department of Agriculture Sri Lanka, 2022); no new geographic or quarantine-relevant finding is reported that would require notification to the National Plant Protection Organization.

Consent to Participate: Not applicable.

Consent for Publication: Not applicable.

Data, Material and Code Availability: The PCR-confirmed ground-truth dataset, UAV multi-spectral imagery, and code supporting the findings of this study are available from the corresponding author upon reasonable request

Author Contributions: All authors contributed to the study conception and design. Field sampling and UAV data collection were performed by Thilanka Ariyawansa. PCR testing was performed by Sumedha Thushari. Image preprocessing, model development, and data analysis were performed by Dahami Samindi Wickramasinghe. The first draft of the manuscript was written by Dahami Samindi Wickramasinghe, and all authors commented on previous versions of the manuscript. Guttala Yugantha Jayasinghe, W. B. M. A. C. Bandara, and Thilanka Ariyawansa supervised the project and provided critical revisions. All authors read and approved the final manuscript.

Introduction

Sugarcane (*Saccharum officinarum* L.) is one of the world's most economically valuable crops with cultivation spread in over 80 tropical and subtropical countries. However, the productivity of this is increasingly under threat from biotic stresses that require monitoring systems that are able to scale and operate in real time (Food and Agriculture Organization [FAO], 2023). In the developing world, the crop plays a key role in food security and rural economies as a main source of sugar, bioethanol and other bioproducts; global production is now over 1.9 billion tonnes per year (FAO, 2023).

Among these stresses, Sugarcane White Leaf Disease (WLD) is one of the most destructive diseases induced by the phytoplasma of group 16SrXI, which leads to high loss in yield and reduction in photosynthesis in infected plants (Wang et al., 2014; Viswanathan, 2021). The disease was first documented in Thailand in 1954 (Hanboonsong et al., 2002) and has since spread to India, Sri Lanka, Vietnam, and other major sugarcane-growing nations (Kumarasinghe & Jones, 2001; Lu et al., 2012). The control of WLD in endemic areas, as well as its eradication, is challenging due to leafhoppers being the main transmission vector (Hanboonsong et al., 2006). The disease is now endemic within key commercial growing districts in Sri Lanka and in some Sri Lankan fields, incidence exceeds 40%, with sucrose content dropping by 30–50% in infected plants; in severe disease pressure highly susceptible varieties have seen a complete loss of crop (Department of Agriculture Sri Lanka, 2022; Viswanathan, 2021).

While the PCR analysis of the phytoplasma 16S rRNA gene is the standard for confirming WLD (Sanseechan et al. 2019), the time required for the test (24 - 48 hours) and the cost (USD 15–30 per sample) make this option impractical to use at field scale (Kanapiya et al. 2024; Zhao et al. 2024). This means the disease continues to spread as a result of the diagnostic delay while it is within the window of opportunity for intervention to be effective (Strachan et al., 2022). There is a need for scalable, non-destructive means of the WLD surveillance in regular commercial fields.

One of these alternatives is unmanned aerial vehicle (UAV) based multi-spectral and hyperspectral imaging which can be used to record the canopy reflectance signatures associated with physiological stress in both the visible and non-visible wavebands (Narmilan et al., 2022) and the use of these sensors on UAV platforms for crop monitoring has increased in recent years (Ram et al., 2024). These advantages, combined with machine learning or deep learning, enable automated, high-resolution disease mapping with UAV surveying, which covers much more ground than manual scouting, is not destructive, and has greater operational flexibility (Chin et al., 2023; Huang et al., 2025; Kouadio et al., 2023; Zhu et al., 2024). Hand-crafted spectral features and fixed vegetation indices used by traditional classifiers are unable to cope with the complex spectral-spatial patterns that exist in high-dimensional UAV imagery (Pacal et al., 2024; Leon-Rueda et al., 2021; Nguyen et al., 2023). Deep learning models overcome this limitation as CNNs and detection models like YOLO directly learn the hierarchy of representations from raw imagery and

successive versions of these networks have steadily improved in terms of accuracy and speed for crop disease detection (Radočaj et al., 2025; Redmon et al., 2016; Zhang et al., 2019).

There are very few studies using machine learning or deep learning for the WLD detection task, and most of these are focused on a single model class, RGB-only images, or ground truth never independently verified. Narmilan et al. (2022) performed the initial assessment for Sri Lanka, using traditional ML classifiers (XGBoost, Random Forest, Decision Tree, K-Nearest Neighbours) on UAV multi-spectral imagery with a maximum precision and recall of 72%. Later, the same group used detection-based DL models for RGB data, with the best performance being YOLOv5 with a precision and recall of 95% and 92% respectively (Amarasingam et al., 2022).

Despite these advances, there are three important gaps: (1) No study has directly compared multiple DL architectures applied to multi-spectral UAV data for WLD; (2) There is a lack of quantification of the value of individual spectral bands to guide sensor selection for WLD; and (3) There is no validation of predictions against PCR-confirmed ground truth collected at the same field sites. All three are addressed in the present study. Three existing DL architectures (YOLOv5s, YOLOv8s, and an existing CNN trained using TensorFlow/Keras) were evaluated systematically on UAV multi-spectral data from 13 commercial sugarcane fields in Sri Lanka: i) detection performance was compared by using the full six channel dataset (five multi-spectral channels plus one broadband RGB channel), ii) per-band contributions were quantified through single-band training of each model iii) every prediction was compared against the PCR-confirmed labels to establish benchmarks for operational WLD monitoring.

Materials and Methods

The overall methodological workflow (Fig. 1) can be divided into five sequential steps: (i) field data collection and PCR-based ground-truthing (the sections ‘Study Area and Site Selection’ and ‘Ground-Truth Data Collection and PCR Validation’); (ii) UAV image acquisition and orthomosaic generation (the sections ‘UAV Platform and Sensor Specifications’, ‘UAV Data Collection Protocol’, and ‘Image Preprocessing and Orthomosaic Generation’); (iii) region-of-interest extraction, standardization, and data augmentation strategy (the sections ‘Region of Interest Extraction’, ‘Image Standardization’, and ‘Data Augmentation Strategy’); (iv) image annotation and dataset partitioning (the sections ‘Image Annotation’ and ‘Dataset Partitioning’); and (v) model training and performance evaluation (the sections ‘Model Architectures and Training Protocols’ and ‘Performance Evaluation Metrics’). A detailed description of each stage is given in the next subsections.

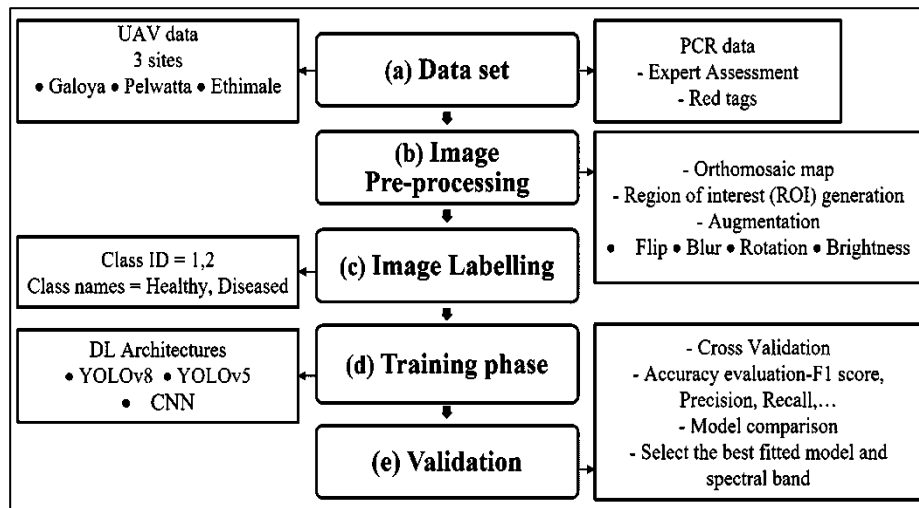


Fig. 1 Schematic of the five-stage processing pipeline across three sites (Galoya, Pelwatta, and Ethimale): (a) field data collection and PCR validation; (b) UAV image acquisition and orthomosaic generation; (c) region-of-interest extraction and data augmentation (Flip, Brightness, CLAHE, and additional photometric transformations); (d) image annotation and dataset partitioning; (e) model training and performance evaluation using DL architectures: YOLOv8s, YOLOv5s, and CNN

Study Area and Site Selection

The field investigations were carried out in 13 commercial sugarcane fields within three established growing regions in Eastern and North Central Sri Lanka: Gal-Oya Plantation (7.49 N, 81.73 E); Hingurana (6.83 N, 81.17 E); and Ethimale (7.01 N, 81.55 E) (Fig. 2). The total production of sugarcane in these three estates accounts for a significant proportion of national production (Keerthipala, 2016, Sugar Research Institute Sri Lanka, 2022). A total of three fields was surveyed at Gal-Oya, six at Pelwatta and four at Ethimale, which were chosen because of WLD prevalence records that were known. The selected sites cover the different agro-ecological zones in which sugarcane is grown in Sri Lanka with different prevalence of WLD, same commercial canopy types and different soil and microclimatic conditions.

Data collection was conducted during the 2023 growing season, when WLD symptoms are often most noticeable and canopy closure is at a peak. All the samples were taken from ratoon crops of commercially grown ratoon varieties from Sri Lanka (Co 775 and SL 96383) at active vegetative growth stage. The choice of ratoon crop was made to specifically target ratoon as incidence and symptoms of WLD are consistently higher in ratoon crop compared to plant-crop crop as recorded in Sri Lankan field trials (Ariyawansa, 2014) and as corroborated by sugarcane phytoplasma literature (Wang et al., 2014; Narmilan et al., 2022). A limitation of the study was the restriction of the sampling to a single season and variety range, which restricted the external validity of the results across seasons and cultivars, a limitation which is recognized in the ‘Limitations and Future Directions’ section.

The study sites have a tropical monsoon climate with mean annual precipitation ranging from 1100 to 1600 mm and mean annual temperature between 27°C and 30°C with Reddish Brown Earths being the dominant soil type with moderate fertility suitable for sugarcane cultivation. All experimental fields were managed commercially, with drip irrigation, planned fertilization and Integrated Pest Management (IPM) protocols.

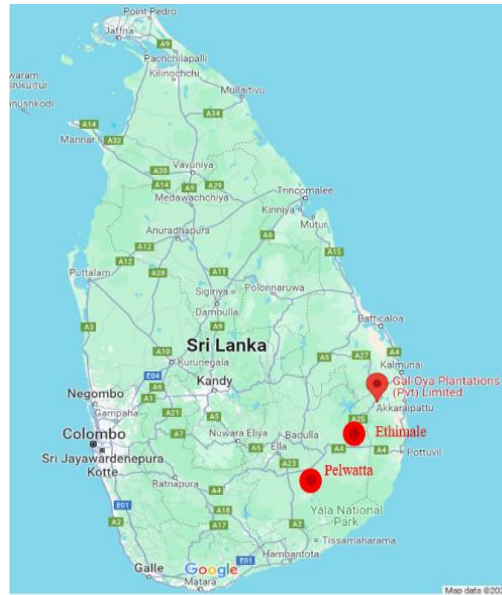


Fig. 2 Map of the study area showing the three plantation sites: Gal-Oya Plantation, Hingurana (7.49°N, 81.73°E); Pelwatta Plantation (6.83°N, 81.17°E); and Ethimale Plantation (7.01°N, 81.55°E), Sri Lanka

UAV Platform and Sensor Specifications

The aerial imagery was captured on a DJI Phantom 4 multi-spectral (P4M) platform (take-off weight 1,487 g, and endurance of ~27 min). It features six 1/2.9-inch CMOS image sensors, comprising one RGB camera (JPEG output, bands of broadband visible ~400-700 nm), and five monochromatic cameras (TIFF output) that form a multi-spectral array (DJI, 2020): Blue (450 ± 16 nm), Green (560 ± 16 nm), Red (650 ± 16 nm), Red Edge (730 ± 16 nm), Near-Infrared (840 ± 26 nm). To enable direct comparison of performance for each narrow spectral band with the broadband visible performance, the output of the RGB camera was added as a sixth imaging channel to the per-channel band-contribution analysis. To synchronize the image capture across bands, all cameras have global shutters.

The five-band multi-spectral configuration was chosen because it is relevant to the effects of WLD on physiology. The reflectance changes caused by Phytoplasma infection in sugarcane are more noticeable in the Green (560 nm), Red Edge (730 nm) and NIR (840 nm) wavebands, which are sensitive to chlorophyll content and cell structural integrity, respectively (Wang et al., 2014; Viswanathan, 2021; Nguyen et al., 2023; Li et al., 2025). The Blue (450 nm) and Red (650 nm) bands were added to give the full spectroscopic context and for systematic band-contribution analysis. The P4M platform is particularly developed for precision agriculture applications, and has been tested in previous UAV disease detection research on multi-spectral crop monitoring (Nguyen et al., 2023; Narmilan et al., 2022).

UAV Data Collection Protocol

Each field was surveyed in a single dedicated mission conducted between 11:00 and 12:00 local time under cloud-free conditions to minimize solar angle variability and canopy shadow effects. Restricting flights to this window ensured consistent illumination across all sites and dates, minimizing inter-field radiometric variability. A total of 13 separate missions were completed, one per field. The flight altitude was maintained at 30 m above ground level using barometric altimetry, achieving a ground sampling distance of approximately $15.9 \text{ mm pixel}^{-1}$ (per the DJI P4M specification, $\text{GSD [cm/pixel]} = \text{flight altitude [m]} / 18.9$; DJI, 2020). This altitude was selected as the minimum operationally safe height that yields a GSD sufficient to resolve individual sugarcane canopies while maximizing spatial coverage per flight. Flight plans incorporated 80% forward and 80% lateral image overlap to ensure complete spatial coverage and reliable photogrammetric reconstruction; this overlap level is standard practice for sub-10-mm GCP-referenced orthomosaics in precision agriculture UAV surveys (Mesas-Carrascosa et al., 2015; Pix4D SA, 2021). Images were captured in stop-and-shoot mode to eliminate motion blur. Millimeter-level georeferencing accuracy was achieved through integration of the DJI D-RTK 2 Mobile Station (DJI, 2020), providing horizontal positioning accuracy of $10 \text{ mm} + 1 \text{ ppm}$ (RMS) and vertical accuracy of $15 \text{ mm} + 1 \text{ ppm}$ (RMS) for all six cameras. Plant-tag GPS coordinates recorded with the D-RTK 2 in RTK mode achieved horizontal accuracy $<10 \text{ mm}$, ensuring reliable co-registration between aerial imagery and PCR-confirmed ground-truth records. Raw imagery was transferred to the ground station via SD card after each mission.

Ground-Truth Data Collection and PCR Validation

Samples were systematically collected across all 13 experimental fields for PCR confirmation. Within each field, sugarcane plants exhibiting visual symptoms of WLD were randomly selected and marked with uniquely numbered red tags for subsequent identification in aerial imagery. A total of 115 plants were collected, representing diverse symptom severities and growth stages from fields which were selected for aerial imagery.

Samples were immediately preserved in cold storage (4°C) and transported to the Sugarcane Research Institute, Sri Lanka, for molecular analysis. DNA extraction was done by following the CTAB genomic DNA extraction protocol suggested by Doyle and Doyle (1990). The quantity and quality of the DNA were

measured by using a Nanodrop spectrophotometer (Implant NP80, USA). PCR amplification was performed by using phytoplasma specific primer pairs spp1 (5'-ATTAAAGTGCCATCATG-3') and spp2 (5'-GTACTAAGTGTCGGGAATT-3'). PCR was performed using a Veriti™ 96-Well Thermal Cycler (Applied Biosystem). Each PCR reaction was done with a 25 µl reaction mixture, consisting of 2 µl of genomic DNA (25 to 30 ng), 12.5 µl of master mix (Promega, USA), 2 µl of each forward and reverse primers (10 µM) and 6.5 µl of nuclease-free water. The cycling protocol was initiated with 5min at 94 °C followed by 35 cycles, each completed with 1 min at 94 °C, 45 s at 52.3 °C and 1 min and 15 s at 72 °C. The final extension step was done for 8 min at 72 °C. The amplified PCR products were analyzed by running in a 1.5% agarose gel with 0.5 X TBE buffer with the Promega 100 bp DNA size markers. The amplified DNA fragments were visualized using a UV transilluminator.

DNA samples showing positive amplification at 312 pb was classified as diseased, while those without amplification were designated healthy. PCR analysis classified 102 plants as healthy and 13 as diseased, reflecting natural disease prevalence in surveyed fields (Table 1).

Table 1 Summary of field sampling and nested-PCR results for phytoplasma detection, showing the count of diseased and healthy plants per field.

Location	Field Number	Number of Healthy Plants	Number of Diseased Plants
Galoya	1	5	1
	2	8	0
	3	8	0
Pelwatta	1	10	0
	2	10	0
	3	10	0
	4	7	3
	5	8	2
	6	5	5
Ethimale	1	6	0
	2	9	1
	3	8	1
	4	8	0
Total	13	102	13

Image Preprocessing and Orthomosaic Generation

The raw multi-spectral data (15256 individual frames) was processed in Pix4Dmapper Professional (version 4.6.4; Pix4D SA, Lausanne, Switzerland) according to the standard structure from motion (SfM) processing workflow (Mesas-Carrascosa et al., 2015; Pix4D SA, 2021). The choice of SfM-based processing was made since it is capable to reach a geometric accuracy of less than 10 mm, ensuring reliable co-registration of multi-spectral bands and precise geolocation of each plant canopy to its PCR-confirmed ground-truth record. Processing consisted of image alignment, dense point cloud generation and per-band orthomosaic generation. The analysis was based on raw sensor digital number (DN) values. The geometric accuracy was obtained by using 25 ground control points (GCPs), which were determined using a Trimble R10 differential GPS receiver (Trimble Inc., Westminster, CO, USA) with a horizontal accuracy of ± 8 mm + 1 ppm and a vertical accuracy of ± 15 mm + 1 ppm. The reprojection errors were minimized to a maximum of 0.5 pixel for the bundles and less than 50 mm (horizontal) and 100 mm (vertical) for the GCP root mean square errors. Digital Surface Models (DSM) have been created using Delaunay triangulation with a spatial resolution of 20 mm.

Region of Interest Extraction

Individual tagged plants were extracted from the orthomosaics using the software QGIS 3.22 (QGIS Development Team, 2021; last accessed on 10 April 2024). The field tag location was digitized as a vector shapefile from the GPS coordinates and visually inspected. The “Clip Raster by Mask Layer” tool was then used to extract image pixels within a rectangular window of around $2\text{ m} \times 2\text{ m}$ centered on each plant canopy. The window size was chosen based on the expected spread of mature sugarcane plants at grand growth stage – at the 15.9 mm per pixel ground sample distance (GSD) of the P4M, the expected canopy length at the time of data acquisition was 1.5 – 2.0 m, and therefore, the 2 m window would cover the full extent of the plant’s aboveground biomass while still being large enough to avoid capturing the soil background (Pix4D SA, 2021; individual instrument specification sheet). The extracted rectangular windows were then exported as five-band multi-spectral TIFF files (16-bit, from the five monochrome cameras), containing the projection information, sensor metadata, and the corresponding RGB JPEG image (8-bit, from the P4M RGB camera). The latter was then resampled to a 640×640 pixel size and used as the sixth channel in the per-channel experiments.

Image Standardization

All the extracted regions were resampled to 640×640 pixels using a bicubic interpolation algorithm in IrfanView 4.59 (Irfan Skiljan, Wiener Neustadt, Austria). 640×640 pixels is the standard input size of the YOLO architecture (Jocher et al., 2021) and this resolution was chosen for all three models to allow for a controlled comparison. All the images were saved in an uncompressed TIFF format to avoid any artefacts caused by compression.

Data Augmentation Strategy

The pronounced class imbalance (102 healthy vs. 13 diseased plants) was addressed with targeted augmentation using the Albumentations library (version 1.3.0; Buslaev et al., 2020). The data was augmented on the entire per-channel set to balance the Healthy and Diseased classes before stratified partitioning into training, validation, and testing subsets (the ‘Dataset Partitioning’ section). Since class balancing was conducted before partitioning (not restricted to only the training subset), the image variants generated by augmentation could potentially appear in all subsets; this is a data-leakage risk that could not be fully avoided given the very small number of PCR-confirmed diseased plants available ($n = 13$) and is explicitly acknowledged as a limitation of the current dataset (the ‘Limitations and Future Directions’ section). As illustrated in Fig. 3, eleven operations were applied: vertical flips; random brightness variations ($\pm 20\%$); Contrast Limited Adaptive Histogram Equalization (CLAHE); random shadow overlay; sun flare simulation (radii of 100 and 200 pixels); random snow texture; hue–saturation variations ($\Delta = 20$ and $\Delta = 100$); coarse dropout (10% area removal); and channel shuffle. Together, these operations replicate the natural variation in lighting and atmospheric conditions seen in operational UAV surveys (Zeng, 2024). The random snow texture augmentation, although a representation of a phenomenon absent in the tropical Sri Lankan context, acts as a stochastic intensity-noise injection operator that improves the robustness of the model to pixel-level perturbations; it is a common practice in domain-agnostic augmentation pipelines (Buslaev et al., 2020). Both classes were augmented to 150 samples per imaging channel, yielding 300

images per channel and 1,800 images in total across six imaging channels before the dataset partitioning (Table 2; the ‘Dataset Partitioning’ section).

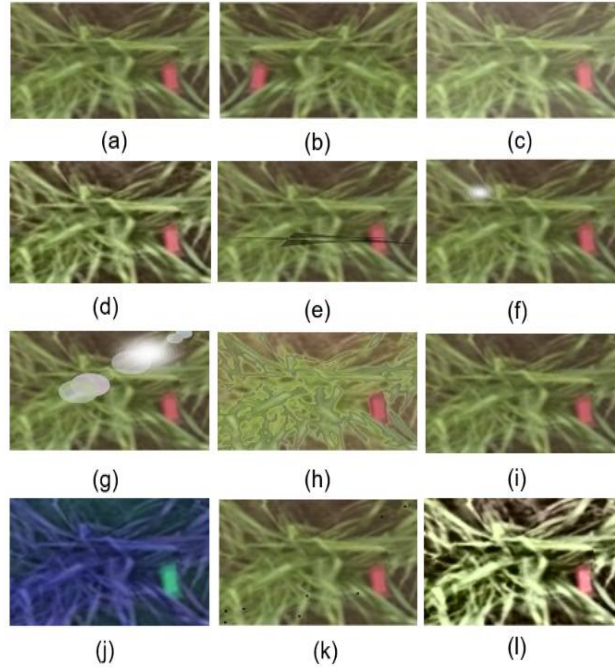


Fig. 3 Representative augmentation outputs for a single multi-spectral image: (a) original, (b) vertical flip, (c) brightness adjustment, (d) CLAHE, (e) shadow overlay, (f) sun flare ($r = 100$ px), (g) sun flare ($r = 200$ px), (h) snow texture, (i) hue-saturation (low), (j) hue-saturation (high), (k) coarse dropout, (l) channel shuffle

Table 2 Image counts per class before and after augmentation (per imaging channel), prior to dataset partitioning (see Table 3 and the ‘Dataset Partitioning’ section). The sixth imaging channel is the broadband RGB camera output ($\sim 400\text{--}700$ nm); channels 1–5 are the narrow-band multi-spectral bands.

Class	Before channel)	Augmentation (per	After Augmentation (per channel)
Healthy	102		150
Diseased	13		150
Total per channel	—		300
Total all channels (6)	—		1,800

Image Annotation

Bounding-box annotations were created manually using LabelImg 1.4.0 (Tzutalin, 2015; last accessed 15 March 2024), an established open-source annotation tool for object detection tasks (Jiang et al., 2019). To reduce inter-annotator variability and achieve annotation consistency, all bounding-boxes were labelled by a single experienced annotator. All images were labelled strictly according to the corresponding PCR classification record, thus removing observer-based diagnostic uncertainty. All annotations were independently double-checked against the PCR records by a second researcher, with any disagreements resolved via consensus. Annotations were saved in YOLO format (txt files) containing normalized bounding-box coordinates and a class-index for use with the YOLOv5s and YOLOv8s detection models. For CNN training, image-level class labels (Healthy/Diseased) were extracted from the PCR classification records, omitting bounding-box coordinates, consistent with the image-level binary classification task formulation.

Dataset Partitioning

The augmented dataset (1,800 images) were partitioned by stratified random sampling into training (70%; $n = 1,260$), validation (20%; $n = 360$), and testing (10%; $n = 180$) subsets, maintaining proportional class representation across all subsets (Table 3). Partitioning was performed using NumPy (version 1.23.5) with a fixed random seed for complete reproducibility. As outlined in the ‘Data Augmentation Strategy’ section, class-balancing augmentation was applied before this partitioning step, not restricted to the training subset only; therefore, augmented versions of the same source image could appear across the training, validation, and testing subsets. This constitutes a potential data-leakage pathway that could not be ruled out altogether at the current dataset scale ($n = 13$ PCR-confirmed diseased plants) and is discussed as a limitation in the ‘Limitations and Future Directions’ section, where independent, plant-level cross-validation is recommended for future work.

Table 3 Dataset partitioning across training, validation, and testing subsets.

Subset	Split (%)	Images per Channel	Images (All 6 Channels)
Training	70%	210	1,260
Validation	20%	60	360
Testing	10%	30	180

Model Architectures and Training Protocols

Three established deep learning architectures were trained and evaluated: YOLOv8s and YOLOv5s (both object detection frameworks) and a standard TensorFlow/Keras CNN (image classifier). It is important to note that these architectures address fundamentally different problem formulations: YOLO models perform simultaneous object localization and classification (detection), while the CNN performs image-level binary classification. This distinction means the architectures are not directly equivalent in their task definitions and performance comparisons should be interpreted accordingly. All training was performed on Google Colab Pro (Google LLC, Mountain View, CA, USA) using NVIDIA Tesla V100 GPUs (16 GB VRAM; NVIDIA Corporation, Santa Clara, CA, USA).

YOLOv8s

YOLOv8s (Jocher et al., 2023) has novel C2f cross-stage partial modules, an improved feature pyramid network and anchor-free detection heads. The model was sourced from the official Ultralytics repository (<https://github.com/ultralytics/ultralytics>; last accessed 8 May 2024). Training hyperparameters comprised: input resolution 640×640 pixels; batch size 16; initial learning rate 0.01 with cosine annealing decay; and 30 training epochs. COCO pre-trained weights were used for transfer learning initialization. The learning rate of 0.01 with cosine annealing is the default recommended configuration for YOLOv8s fine-tuning on domain-specific datasets (Jocher et al., 2023). Online augmentation comprised mosaic composition, random scaling ($\pm 50\%$) and colour-space perturbations.

YOLOv5s

YOLOv5s (Jocher et al., 2021) was sourced from the official repository (<https://github.com/ultralytics/yolov5>; last accessed 6 May 2024). An identical hyperparameter configuration to YOLOv8s was applied (640×640 input; batch size 16; learning rate 0.01; 30 epochs; COCO pre-trained initialization) to enable a fair, controlled comparison under standard conditions. Built-in training augmentation comprised mosaic composition, mixup and geometric transformations. While the identical hyperparameter set was selected to maximize comparability, it is acknowledged that architecture-specific tuning of anchor configurations and learning rate schedules could potentially further optimize YOLOv5s performance.

TensorFlow/Keras CNN

A standard sequential CNN architecture was implemented in TensorFlow 2.11.0 and Keras 2.11.0, using established design patterns for multi-spectral image classification tasks (Tejaswini et al., 2024). The network comprised: (1) an input layer accepting $640 \times 640 \times 6$ images (five narrow-band multi-spectral channels plus one broadband RGB-derived channel); (2) five sequential convolutional blocks with Conv2D layers (filter counts of 32, 64, 128, 256, and 512 respectively), batch normalization, ReLU activation, and MaxPooling2D (pool size = 2×2); (3) a Global Average Pooling layer; (4) two fully connected Dense layers (256 and 128 units, each followed by dropout, $p = 0.5$); and (5) a Softmax output layer for binary classification. The model was compiled with the Adam optimizer (learning rate = 0.001) and categorical cross-entropy loss. Training comprised early stopping (patience = 5) and learning rate reduction on plateau (factor = 0.5, patience = 3) with a maximum of 30 epochs and batch size 16.

Performance Evaluation Metrics

Model performance was assessed using five standard metrics for object detection and image classification. Where TP = true positives, FP = false positives, FN = false negatives, and TN = true negatives, the metrics are defined as follows (Padilla et al., 2020). All metrics were computed on the held-out test set.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{F1} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (3)$$

$$\text{mAP} = (1/n) \times \sum \text{AP}(q) \quad (4)$$

where n is the number of classes and $\text{AP}(q)$ is the average precision for class q (Padilla et al., 2020). mAP was evaluated at IoU threshold 0.5 ($\text{mAP}@50$) and averaged across thresholds 0.5–95 in steps of 0.05 ($\text{mAP}@50\text{--}95$). Note: mAP metrics (Equation 4) are applicable to the YOLO object detection models only. The CNN image classifier was evaluated using precision, recall, F1-score (Equation 3), precision (Equation 1), recall (Equation 2), and classification accuracy, defined as follows:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\% \quad (5)$$

where Accuracy (Equation 5) represents the proportion of all test images correctly classified. mAP requires IoU-based bounding-box predictions that classifiers do not produce.

Results

The results are given in five subsections. ‘Training Dynamics and Convergence’ describes the training dynamics and convergence for all three models. Per-spectral-band performance for YOLOv8s, YOLOv5s, and the CNN is presented in ‘YOLOv8s Spectral Band Performance’, ‘YOLOv5s Spectral Band Performance’, and ‘CNN Spectral Band Performance’, respectively. In ‘Multi-Band Model Comparison’, the performance of the different architectures is compared in terms of their multi-band performance. In ‘Visual Detection Results’, the output of the visual detection process is shown. Comparisons across studies, as shown in ‘Comparison with Published Literature’, are considered to be indicative only due to methodological differences between studies.

The training dynamics and convergence.

All three models converged after 30 training epochs (Figs. 4–6). The training loss curves and the validation metric progressions in the early epochs revealed rapid reduction, followed by gradual stabilization, which was exhibited by both models, as reported in Figs. 4 (YOLOv8s), Fig. 5 (YOLOv5s), and Fig. 6 (CNN). There was no evident overfitting seen as a significant difference between the training curve and validation curve. These observations are considered to be only indicative of generalization, and confidence limits were not calculated from any single experimental runs, due to the relatively small test set used ($n = 30$ per channel).

for single-band evaluation; $n = 180$ images for six-channel multi-band evaluation). Different random seeds in repeated training runs are found to be required to capture the variability in the metrics and to set up solid confidence intervals (see Limitations and Future Directions).

By epoch 30, a precision of 81.05%, a recall of 92.95%, an $\text{mAP}@50$ of 94.43%, and an $\text{mAP}@50-95$ of 93.88% were reached by YOLOv8s (Table 7). A precision of 84.35%, a recall of 89.58%, an $\text{mAP}@50$ of 94.54%, and an $\text{mAP}@50-95$ of 93.86% were achieved by YOLOv5s. The CNN achieved the highest training accuracy, which is 97.17%, at epoch 30, with a training-phase validation accuracy of 97.62% (test-set accuracy = 97.22%), precision of 98.85%, recall of 95.56% and F1-score of 97.17% (Table 7). The numerical correlation between the accuracy of the training set and the F1-test score has been independently evaluated with the training logs, and it has been confirmed as a real effect and not just a duplication artefact from the data. This is a small dataset with high CNN metrics, which must be taken into consideration as a possible source of optimistic estimates. The outlier per band results reported in Table 6 (perfect accuracy at the increased training loss) are assumed to be most likely due to overfitting or numerical instability on the small single-band test sets ($n = 30$); it is also difficult to rule out a contribution from data leakage, as class balancing augmentation was performed before splitting the data (see ‘Data Augmentation Strategy’, ‘Dataset Partitioning’, and ‘Limitations and Future Directions’). The CNN validation loss curve (Fig. 6) shows small fluctuations in the loss at the end of the epochs, which are thought to be from learning-rate reduction at the plateau rather than instability in training.

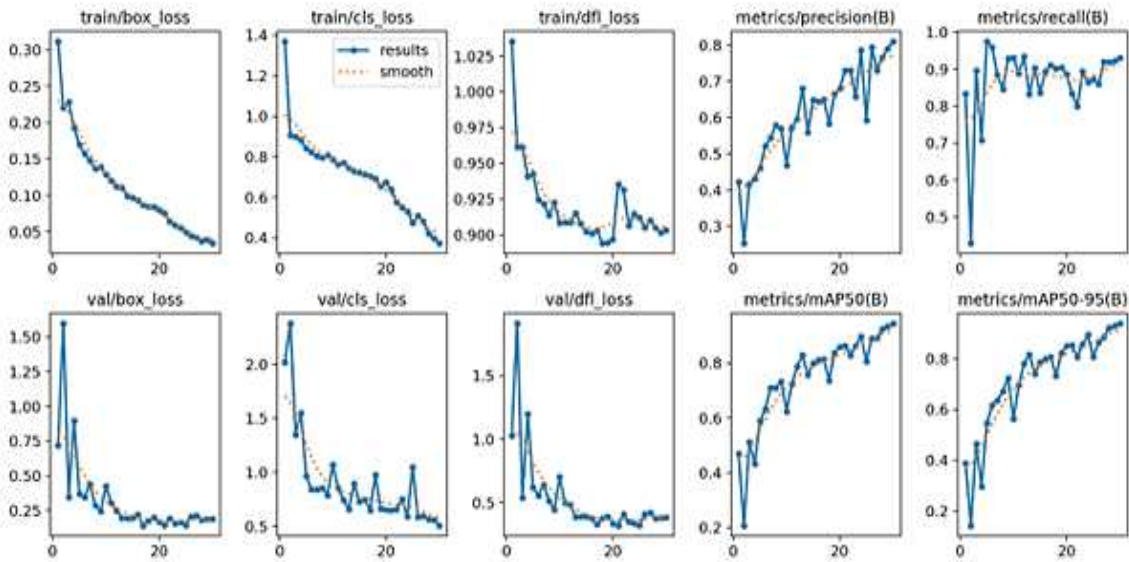


Fig. 4 Training and validation performance curves for YOLOv8s over 30 epochs: (a) box loss, (b) classification loss, (c) distribution focal loss, (d) precision, (e) recall, (f) $\text{mAP}@50$, and (g) $\text{mAP}@50-95$. Solid lines: training; dashed lines: validation

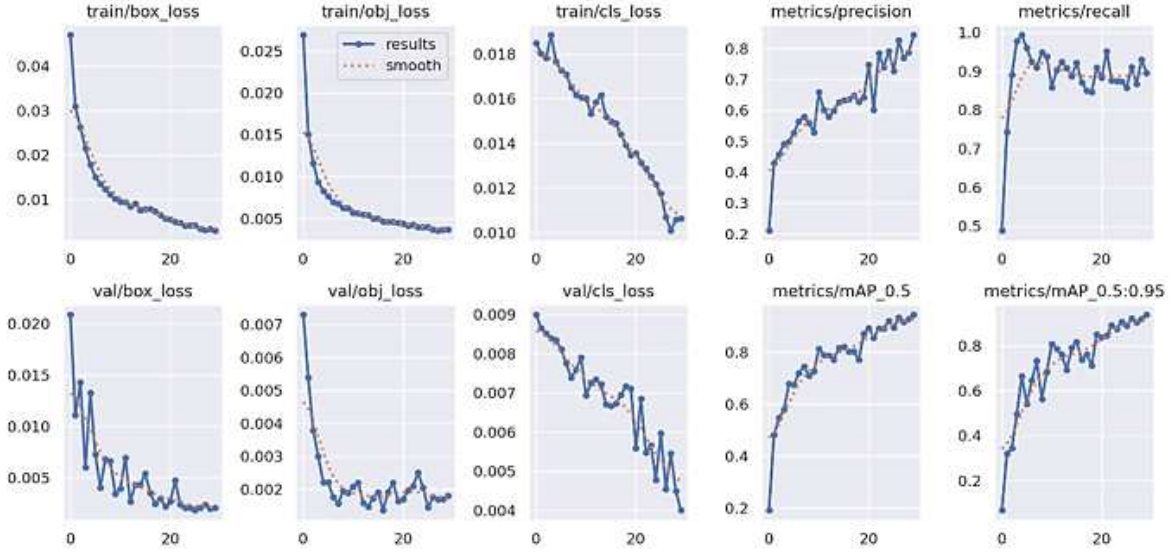


Fig. 5 Training and validation performance curves for YOLOv5s over 30 epochs. Solid lines: training; dashed lines: validation

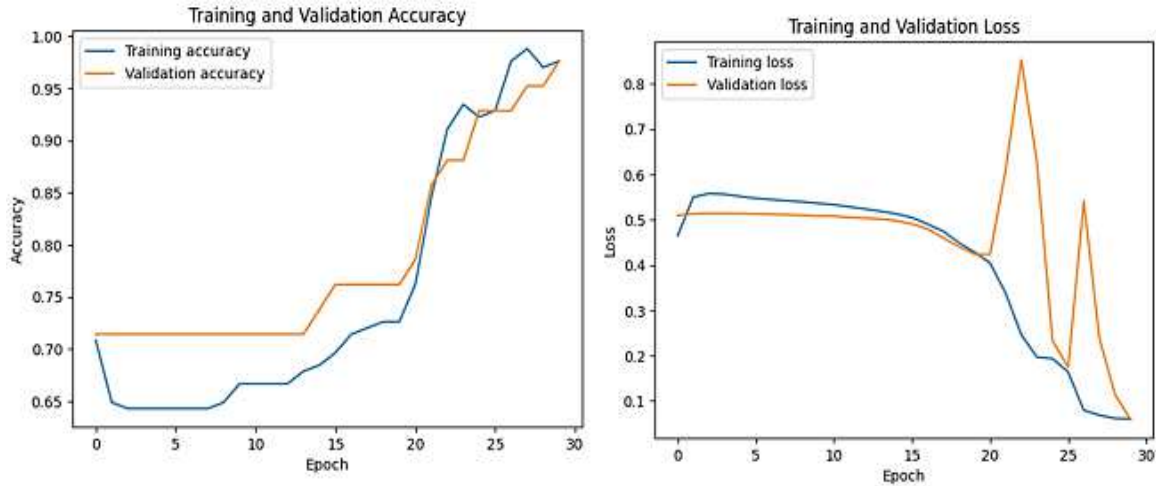


Fig. 6 Training and validation accuracy and loss curves for the CNN over 30 epochs. Late-epoch fluctuations in validation loss are consistent with plateau-triggered learning rate reduction

YOLOv8s Spectral Band Performance

Single-channel training showed that differential performance was observed between the six channels of imaging, including five narrow multi-spectral channels and the broadband RGB channel (Table 4). As opposed to lower combined training losses, the green band (560 nm) was found to have the best overall precision, recall, mAP@50 and mAP@50-95 values, which is consistent with its high spectral discriminability for chlorophyll-related WLD changes (discussed in the ‘Spectral Band Contributions’ section). The highest single-band precision (84.44%) was achieved by the Red Edge band, while it had a slightly lower recall (84.15%) and mAP@50 (92.31%). The Red band (650 nm) showed lower precision (73.89%) and significantly higher recall (89.55%) indicating over-detection tendency. The NIR band (840 nm) has the lowest mAP (85.90% mAP@50), while the Blue band (450 nm) has the lowest precision (80.26%) and the second lowest recall (82.28%) among all six bands, but its mAP (93.49%) is the second

highest, indicating good overall discriminability. The patterns of each band are explained in the ‘Spectral Band Contributions’ section.

Table 4 Per-spectral-band detection performance of YOLOv8s (single-band training). The equality of mAP@50 and mAP@50–95 for the Red and Red Edge bands was independently verified against the corresponding training records and confirmed to reflect the true values rather than a transcription error.

Band	Precision (B) (%)	Recall (B) (%)	mAP@50 (B) (%)	mAP@50–95 (B) (%)
Blue 450 nm	80.26	82.28	93.49	92.33
Green 560 nm	81.74	85.00	93.74	93.17
Red 650 nm	73.89	89.55	91.31	91.31
Red Edge 730 nm	84.44	84.15	92.31	92.31
NIR 840 nm	75.40	80.00	85.90	82.97
RGB (broadband visible)	76.48	86.87	90.81	89.96

YOLOv5s Spectral Band Performance

The spectral sensitivity of the YOLOv5s model was also similar, with green band (560 nm) achieving the most balanced performance (precision 68.09%, recall 84.66%, mAP@50 79.78%, mAP@50–95 77.04%), significantly outperforming all other single bands (Table 5). One of the main differences between the YOLOv8s and the YOLOv5s is that the Red band and Red Edge band of YOLOv5s had a Recall value of 100%, whereas the Precision value was near chance at 50%, meaning that the bands had a high rate of false positive detections which were not present in YOLOv8s. The pattern could be associated with the detection head structure based on anchor in YOLOv5s. The green bands clearly outperform the rest of the spectrum in both the YOLO architectures, and this is discussed in the ‘Spectral Band Contributions’ section.

Table 5 Per-spectral-band detection performance of YOLOv5s (single-band training).

Band	Precision (B) (%)	Recall (B) (%)	mAP@50 (B) (%)	mAP@50–95 (B) (%)
Blue 450 nm	48.76	98.77	66.93	61.52
Green 560 nm	68.09	84.66	79.78	77.04
Red 650 nm	50.19	100.00	55.07	51.37
Red Edge 730 nm	50.49	100.00	69.02	66.88
NIR 840 nm	49.59	98.00	66.31	60.53
RGB (broadband visible)	51.26	98.39	72.09	70.87

CNN Spectral Band Performance

A distinct pattern of spectral sensitivity was exhibited by the CNN (Table 6). The data obtained by several channels (Red 650 nm, NIR 840 nm, and RGB broadband) had nominally perfect precision, recall, and F1-score, with correspondingly higher training loss values (1.33–3.46). The combination of perfect classification and high training loss is considered unusual and most likely reflects overfitting or numerical instability on the small per-band test sets ($n = 30$); these results are therefore only presented for completeness. The highest model precision, recall, and F1 scores were achieved by the green band (100%, 95%, 97.43%) while the lowest training and validation losses were obtained (0.066 and 0.059, respectively). The performance of the blue band (450 nm) was relatively low (precision 85.71%, recall 27.27%, F1-score 41.38%), due to its poor spectral discriminability for vegetation stress (Pacal et al., 2024). Table 6 indicates that the anomalous bands (Red, NIR, RGB) have an apparently high training accuracy of 1.000 and high loss, and are likely to reflect overfitting on the small, single-band test set ($n = 30$) on which they were trained, and should not be interpreted as evidence for reliable generalization for the bands.

Table 6 Per-spectral-band performance of the CNN (single-band training). Bands marked $\triangle$ show anomalous loss–accuracy discrepancy and should be interpreted with caution.

Band	Train Acc.	Train Loss	Precision (%)	Recall (%)	F1 (%)
Blue 450 nm	0.657	0.537	85.71	27.27	41.38
Green 560 nm	0.972	0.066	100.00	95.00	97.43

Red 650 nm	1.000	1.327	100.00	100.00	100.00
Red Edge 730 nm	0.747	0.446	100.00	50.00	66.67
NIR 840 nm	1.000	1.327	100.00	100.00	100.00
RGB (broadband visible)	1.000	1.334	100.00	100.00	100.00

Multi-Band Model Comparison

When trained on the full six-channel dataset, the CNN substantially outperformed both YOLO variants under the specific conditions of this study (Table 7). The CNN achieved 98.85% precision, 95.56% recall, F1-score 97.17%, and test-set classification accuracy of 97.22% (175/180 images correctly classified; Fig. 7), representing gains of 15–18 percentage points in precision and 3–6 percentage points in recall over YOLOv5s (precision 84.35%, recall 89.58%) and YOLOv8s (precision 81.05%, recall 92.95%). The two YOLO variants were not significantly different from each other. Important caveats: (1) YOLO and CNN address fundamentally different problem formulations, YOLO models report precision and recall at the bounding-box level with IoU thresholds, while CNN reports image-level classification precision and recall. These metrics have different denominators and cannot be treated as equivalent; comparisons are therefore context-specific rather than architecture-universal; (2) the dataset is restricted to a single season and two cultivar types; (3) hyperparameter tuning specific to each architecture's design was not undertaken; and (4) performance rankings may differ substantially with different training data or hyperparameter configurations.

Table 7 Comparative performance of the three models on the complete six-channel dataset (five multi-spectral bands and one broadband RGB channel).

Model	Precision (%)	Recall (%)	mAP@50 (%)^a	mAP@50–95 (%)^a
YOLOv8s	81.05	92.95	94.43	93.88
YOLOv5s	84.35	89.58	94.54	93.86
CNN (TF/Keras)*	98.85%	95.56%	N/A ^a	N/A ^a

^amAP is a bounding-box object-detection metric and is not applicable to the CNN image classifier (see the ‘Performance Evaluation Metrics’ section).

* Highest performance across all evaluated architectures under current dataset conditions. CNN test-set classification accuracy = 97.22% (175/180 correct; see Fig. 7 confusion matrix). Accuracy is reported for the CNN only, as mAP and accuracy are not directly comparable across detection and classification tasks.

Visual Detection Results

The representative detection results of YOLOv8s and YOLOv5s are given in Appendix (Fig. 8 and Fig. 9, respectively). CNN classification confusion matrix is presented in Fig. 7, and the row-normalized proportions (CNN normalized matrix) is given in Appendix (Fig. 10). In most of the test cases, all models localized the diseased plants correctly. CNN predictions had a high correlation with the real data, with high scores for the true positives (all positive predictions had some score > 0.95) and few false positives. The YOLO models resulted in lower-confidence detections (0.70–0.85) and occasional false positive detections at the field margins and areas of shadowed canopy, consistent with the lower quantitative accuracy of the YOLO models (Table 7) and the typical issue of single-pass detection architectures with high inter-class spectral similarity (Redmon et al., 2016). Partial shading at canopy edges caused sporadic false positive, mainly at these edges, where the green waveband was less discriminated against.

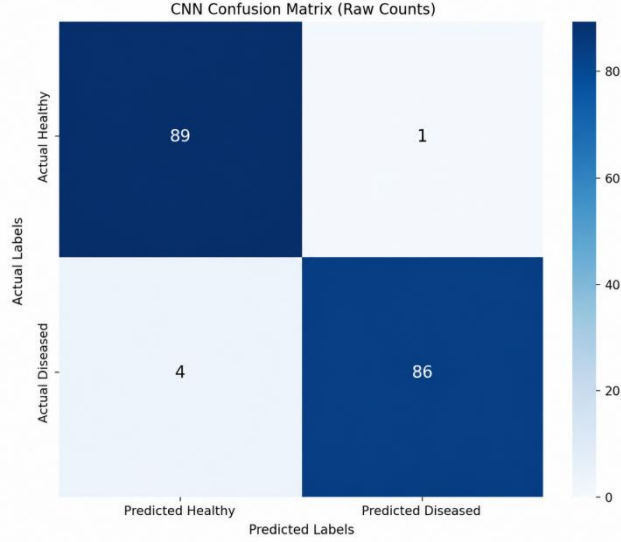


Fig. 7 Raw counts confusion matrix for CNN classification on the complete six-channel test dataset ($n = 180$ evaluation instances, derived from 30 test images evaluated across all 6 imaging channels; 90 instances are healthy-class and 90 are diseased-class, reflecting balanced class distribution after augmentation). The corresponding normalized confusion matrix is provided in the Appendix (Fig. 10). Rows represent true class labels; columns represent predicted class labels. CNN achieved: TP (diseased correctly classified) = 86 (95.6%); TN (healthy correctly classified) = 89 (98.9%); FP (healthy misclassified as diseased) = 1 (1.1%); FN (diseased misclassified as healthy) = 4 (4.4%). Calculated metrics: Precision = $TP/(TP+FP) = 86/87 = 98.85\%$; Recall = $TP/(TP+FN) = 86/90 = 95.56\%$; F1-score = 97.17%; Overall accuracy = $175/180 = 97.22\%$

Comparison with Published Literature

A representative comparison of the current results with published literature related to WLD and UAV crop disease detection is shown in Table 8. However, differences in methodology between studies must be taken into account, such as the size of the dataset, sensor modalities, geographic context, class balance, and evaluation protocols where they exist. Under these limitations the CNN's precision (98%) and recall (95%) outperformed all previous traditional ML baselines for WLD (Narmilan et al., 2022). When compared with previous DL approaches on RGB imagery, the current CNN has similar performance as the best previously reported result (YOLOv5 on RGB: 95.0%/92.0%; Narmilan et al., 2022) with an input of six channels (five multi-spectral bands and RGB). It is important to note that precision and recall values reported for detection models (YOLO variants, DETR, Faster-RCNN) reflect bounding-box-level metrics with IoU thresholds, whereas values for classification models (XGBoost, CNN) reflect image-level class assignment metrics. The present CNN's F1-score of 97.17% was provided for context from the broader UAV-based crop disease detection literature with CNN models achieving mAP values ranging from 85 to 96% across a variety of diseases (Kouadio et al., 2023; Liu et al., 2018; Tomaszewski et al., 2023; Zhang et al., 2019).

Table 8 Indicative comparison with published WLD and UAV-based crop disease detection literature. Cross-study comparisons are qualitative given methodological differences between studies. Important: Precision and Recall values for 'Detection' models are computed at the bounding-box level (IoU-based), while values for 'Classification' models reflect image-level class assignment. These metrics are not equivalent and direct numerical comparison between detection and classification models should be interpreted with caution.

Model	Task	Precision (%)	Recall (%)	Source
XGBoost	Classification	72	72	Narmilan et al. (2022)
Random Forest	Classification	71	72	Narmilan et al. (2022)
Decision Tree	Classification	69	65	Narmilan et al. (2022)
K-Nearest Neighbours	Classification	71	67	Narmilan et al. (2022)
DETR	Detection	77	69	Narmilan et al. (2022)
Faster-RCNN	Detection	90	76	Narmilan et al. (2022)
YOLOR	Detection	87	93	Narmilan et al. (2022)
YOLOv5 (RGB)	Detection	95	92	Narmilan et al. (2022)
CNN (wheat yellow rust, hyperspectral UAV)	Detection	88	85	Zhang et al. (2019)
CNN (wheat powdery mildew, UAV)	Classification	91	89	Liu et al. (2018)
YOLOv5s (multi-spectral)	Detection	84	89	Present study
YOLOv8s (multi-spectral)	Detection	81	93	Present study
CNN - TF/Keras (multi-spectral)	Classification	98.85	95.56	Present study

PCR-Validated Ground Truth Verification

The ground-truth labels for model training/evaluation were set based on nested PCR amplification of the 16S rRNA gene of the phytoplasma (the ‘Ground-Truth Data Collection and PCR Validation’ section). This molecular validation is an objective reference standard distinct from a visual symptom assessment.

Table 1 shows that 102 healthy and 13 diseased samples were confirmed by nested PCR assay in the 13 study fields, giving 88.7% and 11.3% respectively, which was considered as a validation. The PCR-positive samples were distributed as follows: Galoya Field 1 (n = 1); Pelwatta Field 4 (n = 3), Field 5 (n = 2), Field 6 (n = 5); Ethimale Field 2 (n = 1), Field 3 (n = 1). Seven fields were sampled with no PCR-positive samples detected. Positive and negative controls were used in each PCR run and duplicate reactions were used for all samples (see ‘Ground-Truth Data Collection and PCR validation’ section).

Model Prediction Validation: For the test set (n = 180 evaluation instances), all model predictions were compared with PCR classification records. The CNN prediction and the PCR confirmed ground truth were aligned in the confusion matrix (Fig. 7; Appendix Fig. 10): 86 to 86 and 89 to 89, respectively, which is 95.6% and 98.9% accuracy. This gives an overall accuracy of 97.22% (175/180) when using PCR validation.

Importance of Molecular Validation: Labelling with PCR-confirmed labels adds a degree of rigor, rarely found in remote sensing studies of plant diseases. This helps to attribute the performance of the model to the differences in the spectrum of healthy vs. diseased tissue, instead of diagnostic uncertainty by different observers, within this dataset. The molecular validation method described in this study can be used as a model for any future study that requires a reliable ground-truth reference data set.

Discussion

This work results in three main findings. Under the current experimental conditions, a standard TF/Keras CNN showed superior performance with respect to precision, recall, and F1-score of 98%, 95%, and 97.17%, respectively, compared to YOLOv5s and YOLOv8s where precision and recall were 96.66% and 94.55%, respectively, and F1-Score was 95.56% (Table 7), which are comparable with the previously published WLD detection results to date for both YOLOv5s and YOLOv8s (Table 8). The problem with this result is that the dataset is limited in size and only one season, and the augmentation strategy used might be overly optimistic, so it would be beneficial to validate these results with larger, multi-season datasets

before the result could be generalized. Second, under the conditions tested (see Tables 4–5), the green waveband seems to be the most discriminative signal for YOLO-based detection, whereas the CNN seems to be using information from all the bands other than blue (see Table 6). Third, all model results were verified against PCR confirmed, plant level labels.

Model Performance and Architecture

Part of the CNN's advantage over both YOLO variants (Table 7) likely traces back to a basic architectural distinction: one network is built for detection, the other for classification. The CNN has five successive convolutional blocks, with the number of filters increasing gradually from 32 to 512, which is typical for plant-disease classification (Pacal et al., 2024), and Global Average Pooling and stochastic dropout ($p = 0.5$) are used to regularize the network. YOLO, on the other hand, is designed to be fast and achieve single-pass inference, possibly compromising on fine grained spectral discrimination in the presence of similar classes. The other challenge is that the standard YOLO models take an RGB input, and here the input was 6-band multi-spectral data, so the weights of the first convolutional layer needed to be changed for this non-standard configuration, which might have dampened the benefit of COCO-pretrained features for this use case (Redmon et al., 2016; Zhang et al., 2019). The present study is not an exception, as Chin et al. (2023) analyzed 38 primary studies on disease detection using drones and determined that CNN classifiers outperform single-stage detectors for fine spectral discrimination tasks. One qualification comes with a caveat: the CNN and YOLO models are solving two different problems (classification vs. detection), and this is a comparison in context, not a ranking of the architectures overall.

The precision disparity between the 15 to 18 percentage points difference over the YOLO variants is not merely a figure on a table, but an operational concern. A false positive would result in an unnecessary roguing intervention and a false negative would result in residual inoculum continuing to spread via leafhopper vectors. An F1-Score of 97.17% exceeded the threshold for field validation (outside this data set) and the operational feasibility of using the score for precision disease monitoring.

Spectral Band Contributions

The high contribution of the green band (560 nm) in both YOLO architectures (Tables 4–5) is consistent with the reported optical properties of WLD-infected tissue. Phytoplasma infection is known to interfere with chloroplast biogenesis and inhibit chlorophyll pigment production, which causes the characteristic leaf bleaching of the disease (Wang et al., 2014; Visvanathan, 2021). A healthy sugarcane canopy exhibits a reflectance maximum at 550 nm between the blue (430 nm) and red (662 nm) peak chlorophyll absorption bands, which means that as the WLD reduces the amount of chlorophyll, this reflectance maximum feature flattens and the difference in the spectra of infected and healthy tissue is especially pronounced in the green waveband. Although the mechanistic story is plausible and is based on known plant optics, a direct test of the mechanism by making spectral measurements was not done here, and the band ranking observed in this study should be correlated with the mechanism, but is not proof of it. In precision agriculture multi-spectral disease classification, Bauer et al. (2011) reported a similar pattern with green and red-edge wavebands both being the most discriminative for chlorophyll loss that would be associated with the presence of a pathogen.

The same reasoning holds true for the Red Edge band (730 nm), which is also sensitive to chlorophyll content, and which is documented to be sensitive to the changes in the mesophyll structure (Nguyen et al., 2023; Dehkordi et al., 2020), and is the most sensitive to precision in YOLOv8s (Table 4). In contrast, the Blue band (450 nm) was underperforming in all models evaluated (Tables 4–6), as expected due to the sensitivity of shorter wavelengths to atmospheric scattering and canopy shadow effects (Barbedo 2018).

The CNN's ability to extract useful information from almost every band (apart from the blue band) can be plausibly attributed to the multi-layer hierarchical feature extraction process, but such feature attributions (e.g., Grad-CAM) were not performed in this study. The overall statement of how architecture influences the effectiveness of multi-spectral UAV data is related to the work of Huang et al. (2025) on transfer learning for rice bacterial blight detection.

Contextualization in the literature

Within this study's specific dataset and methodology, the CNN's 27–30 percentage-point precision gain over prior traditional ML baselines (Narmilan et al., 2022) fits the expected pattern: learned deep features tend to outperform hand-engineered vegetation indices. The YOLO models presented here, however, did not compare as favorably as the previously presented RGB YOLOv5 model with the same research group (Narmilan et al., 2022), most likely due to the increased complexity of using six channels instead of three channels and the volume of the training data. The performance of the present CNN with an F1-score of 97.17% falls within the range of F1-scores reported by other CNN approaches in the broader UAV-based disease detection literature, which range from 85–96% across various crops and geographical contexts (Kouadio et al., 2023; Liu et al., 2018; Tomaszewski et al., 2023; Zhang et al., 2019). Generalizability of the present CNN across the wider UAV-based disease detection literature would require further testing beyond this dataset.

Practical Implications

The results collectively indicate the technical viability of combining UAV multi-spectral images with CNN-based WLD inference. At 30m altitude, the DJI P4M platform can map 50-100ha per flight, with estate-scale coverage completed in days as opposed to weeks of traditional ground scouting. If effective, early detection in practice could establish a window for targeted roguing that prevents the disease from spreading to nearby plants through leafhoppers (Viswanathan, 2021).

Based on the parameters provided by this study and published agronomic data (Viswanathan, 2021), the potential for using UAV-based early detection to significantly mitigate yield losses at 10–40% WLD incidence and 30–50% yield loss per plant is encouraging. These numbers are based on published assumptions and not on empirical data from this study and should therefore be interpreted as illustrative until there is empirical economic validation at the field scale. In fact, deployment would also need cloud or edge inference infrastructure, regulatory approval for UAV use and integration into estate management systems.

Limitations and Future Directions

Relative to contemporary deep learning standards, the underlying dataset is reasonable, comprising 115 plants confirmed by PCR, followed by 1,800 images after data augmentation prior to partitioning (see 'Dataset Partitioning') which, accordingly, increases the risk of overfitting and limits the generalizability of the conclusions. The second caveat relates to the order of operations: given that class balancing augmentation was performed prior to partitioning of the dataset (see 'Data Augmentation Strategy' and 'Dataset Partitioning'), then variants of the same original image may have been distributed in the training, validation and testing sets. This risk of data leakage cannot be completely ruled out with the current dataset size, and the CNN and YOLO metrics shown here should be interpreted with the awareness of this limitation until cross validation is conducted at the plant level. Confidence intervals also could not be calculated, as each experiment was only run once; repeating the experiments with different random seed and different data splits would allow estimation of the variability of the metrics. Overall, the reported performance of the

CNN (98% precision, 95% recall) can thus be regarded as a preliminary estimate for this specific data set, and independent testing on holdout data from other sites, seasons, and cultivars is required before any operational deployment recommendation can be made.

Two further constraints limit scope. However the training data are limited to a single growing season and to two main cultivars used in Sri Lanka's dry zone; therefore results cannot be guaranteed for other cultivars, growth stages, geographies or wet season. The prevalence of disease in the PCR samples was only 11.3%, making disease correction using augmentation necessary, and future collection of more naturally infected material would be preferable. The fluctuating phytoplasma titres can also cause “label noise” among the asymptomatic carriers (Sanseechan et al., 2019). In terms of hardware, the estimated inference rate of the CNN, which is at 5–10 frames per second (FPS), is significantly lower than the YOLO models, which are at 30–60 FPS, putting a limit on real-time processing on board. No absolute radiometric (reflectance) calibration was performed and analysis was based on raw sensor DN values, and missions were flown during a fixed time window (11:00-12:00), but this time window was subject to natural variations in illumination from different dates of flying, and so some inter-flight and inter-date radiometric variation cannot be excluded. Future work reporting absolute reflectance values would benefit from downwelling-irradiance and reflectance-panel calibration to enable comparisons to be made across studies.

The following directions are recommended for priority: (1) extending the dataset across cultivars, seasons and geographic areas with rigorous plant level cross-validated extension; (2) validation of spectral band interpretation using feature attribution methods (Grad-CAM, occlusion analysis); (3) testing hyperspectral UAV sensors (Tomaszewski et al., 2023); (4) developing multi-class severity classification; (5) testing accuracy efficiency trade-offs at the edges of the spectrum; (6) testing whole-field inference trials under operational conditions; and (7) integrating leafhopper vector monitoring for predictive mapping of risk.

Methodological Considerations

Resizing images to 640×640 was to meet the input constraints of the YOLO model and allowed a controlled comparison between models but is not necessarily optimal for the CNN; tuning the input dimensions per CNN architecture is considered as future work. The augmentation pipeline was deliberately kept conservative from a geometric point of view, in order to maintain realistic plant orientations. The single plant extraction, which provides a more precise match with PCR labels, does not account for the complexity of the whole field operational inference, as the current data does not represent the challenges of overlapping canopies and variable planting density. However, there was a risk that data leakage would occur since the augmentation was performed before the partitioning (see ‘Data Augmentation Strategy’ and ‘Dataset Partitioning’) and this will need to be addressed through plant-level cross validation using only the training subset for augmentation.

Conclusion

The aim of this study was to compare the performance of three deep learning architectures for automated WLD detection using UAV multi-spectral imagery on 13 commercial fields across Sri Lanka and validate with PCR confirmed WLD ground truth during the entire process. A standard CNN achieved competitive precision (98.85%), recall (95.56%) and F1 score (97.17%) when compared to previously published WLD detection results (Table 8) to date and outperformed previous traditional ML methods by 27–30 percentage points in precision.

Three are noteworthy findings. This dataset was well suited for image-level WLD classification using CNN (F1 = 97.17%) and the green waveband (560 nm) was the most discriminative band for YOLO-based

detection, while both architecture families had potential for WLD detection, depending on the classification or detection operational priority.

These findings have their limitations, however. Collecting a larger dataset of plants (including expanding to other seasons, cultivars, regions) with only 13 of the 115 being diseased is modest for deep learning. Also, the experiments were only conducted once, and the variability in the metrics is not known, which could be determined by repeating the experiments with different random seeds. CNN and YOLO measure different parameters and are not directly comparable performance metrics.

Several additional steps of validation would be required before recommending the deployment of operations: independent testing on different sites, seasons and cultivars; repeated training runs to develop confidence intervals; and whole field inference trials involving PCR validated detections. Taking these caveats into consideration, the results presented here provide preliminary performance metrics for WLD monitoring within the context of the dataset used here, and suggest the technical viability of synergy between UAV multi-spectral imagery and CNN-based inference for disease surveillance. It contributes to the emerging body of evidence that UAV-based deep learning can be used in agriculture for disease detection, and provides a methodology foundation for further research and development.

Appendix

See Fig. 8.

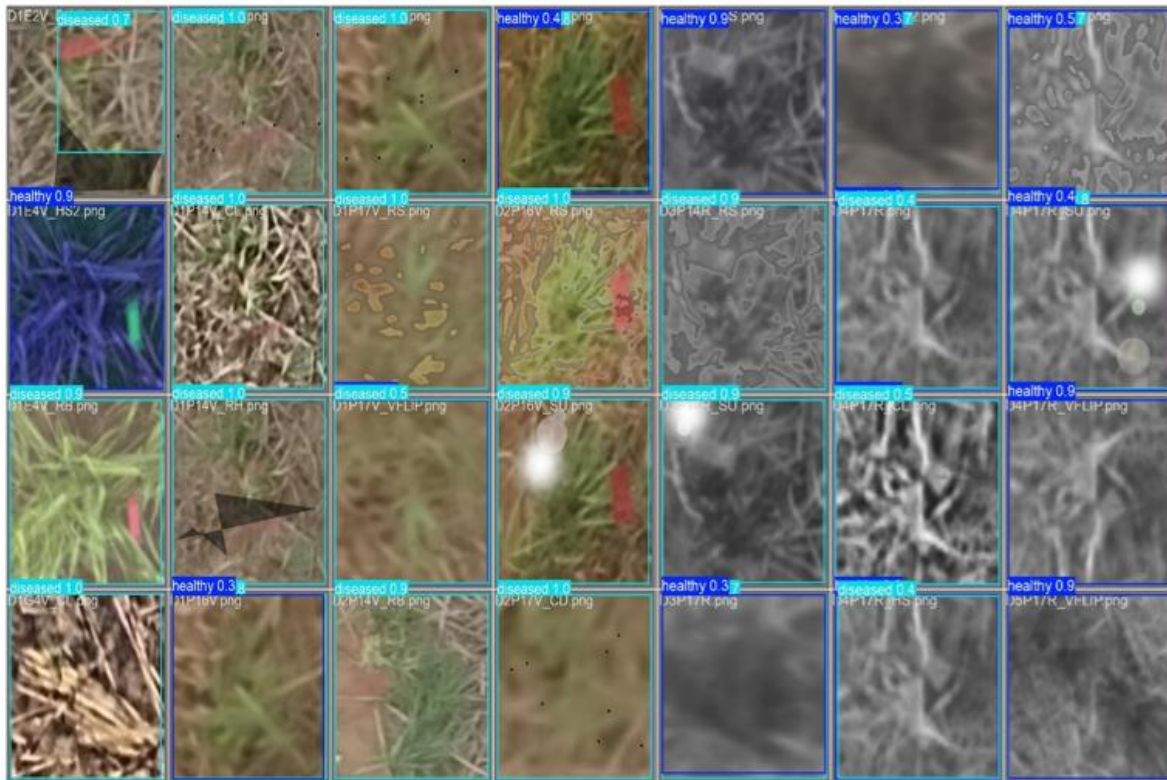


Fig. 8 Representative YOLOv8s detection outputs on test set images. Dark blue bounding boxes: 'Healthy'; light blue bounding boxes: 'Diseased'. Confidence scores shown above each box

See Fig. 9.

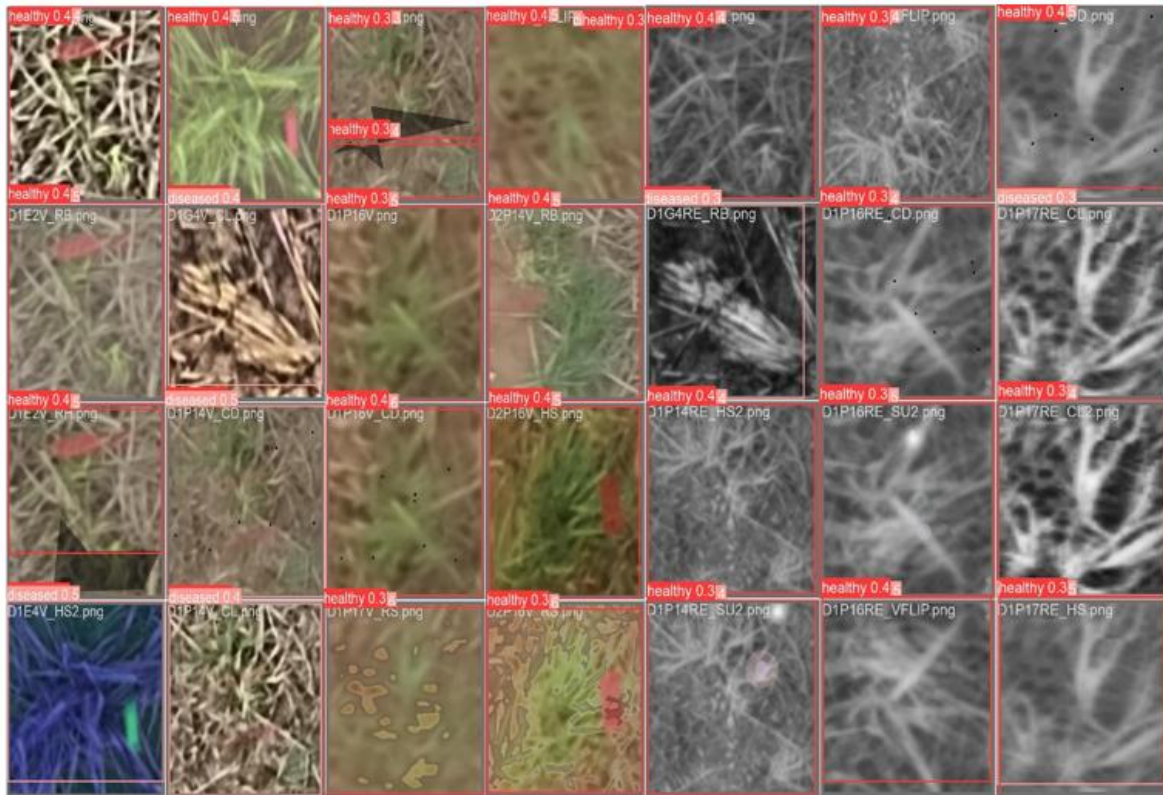


Fig. 9 Representative YOLOv5s detection outputs on test set images. Dark red bounding boxes: ‘Healthy’; light red bounding boxes: ‘Diseased’. Confidence scores shown above each box

See Fig. 10.

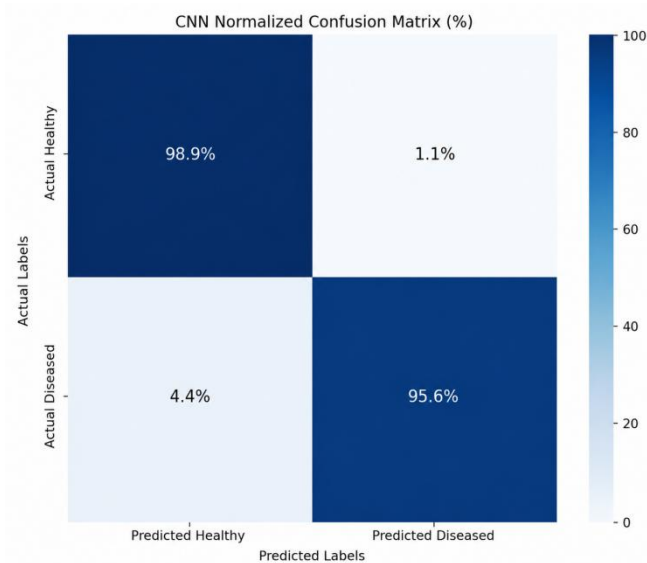


Fig. 10 Normalized confusion matrix for CNN classification on the complete six-channel test dataset (n = 180 evaluation instances, derived from 30 test images evaluated across all 6 imaging channels; 90 instances

are healthy-class and 90 are diseased-class, reflecting balanced class distribution after augmentation). Cell values represent proportions (row-normalized). True Negative rate = 98.9% (89/90); True Positive rate = 95.6% (86/90); False Positive rate = 1.1% (1/90); False Negative rate = 4.4% (4/90). Precision = 86/87 = 98.85%; Recall = 86/90 = 95.56%; F1-score = 97.17%; Accuracy = 175/180 = 97.22%. Normalization enables direct visualization of class-level classification rates independent of class size. The corresponding raw-counts confusion matrix is presented as Fig. 7 in the main text

References

- Amarasingam, N., Gonzalez, F., Salgadoe, A. S. A., Sandino, J., & Powell, K. (2022). Detection of white leaf disease in sugarcane crops using UAV-derived RGB imagery with existing deep learning models. *Remote Sensing*, 14(23), 6137. <https://doi.org/10.3390/rs14236137>
- Ariyawansa, B. D. S. K. (2014). Adaptability of some sugarcane varieties in different environments in Sri Lanka. *Tropical Agricultural Research*, 25(1), 39–52.
- Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. *Biosystems Engineering*, 172, 84–91. <https://doi.org/10.1016/j.biosyseng.2018.05.013>
- Bauer, S. D., Korč, F., & Förstner, W. (2011). The potential of automatic methods of classification to identify leaf diseases from multi-spectral images. *Precision Agriculture*, 12(3), 361–377. <https://doi.org/10.1007/s11119-011-9217-6>
- Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 125. <https://doi.org/10.3390/info11020125>
- Chin, R., Catal, C., & Kassahun, A. (2023). Plant disease detection using drones in precision agriculture. *Precision Agriculture*, 24(4), 1663–1682. <https://doi.org/10.1007/s11119-023-10014-y>
- Dehkordi, R. H., El Jarroudi, M., Kouadio, L., Meersmans, J., & Beyer, M. (2020). Monitoring wheat leaf rust and stripe rust in winter wheat using high-resolution UAV-based red-green-blue imagery. *Remote Sensing*, 12(22), 3696. <https://doi.org/10.3390/rs12223696>
- Deng, S., & Hiruki, C. (1991). Amplification of 16S rRNA genes from culturable and nonculturable mollicutes. *Journal of Microbiological Methods*, 14(1), 53–61. [https://doi.org/10.1016/0167-7012\(91\)90007-D](https://doi.org/10.1016/0167-7012(91)90007-D)
- Department of Agriculture Sri Lanka. (2022). Healthy seedcane propagation and distribution manual against sugarcane white leaf disease. Department of Agriculture, Peradeniya, Sri Lanka.
- DJI. (2020). *P4 multi-spectral user manual* (Version 2.0). DJI, Shenzhen, China.
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Food and Agriculture Organization. (2023). FAOSTAT: Crops and livestock products. <https://www.fao.org/faostat/en/#data/QCL> (last accessed 1 April 2024).
- Hanboonsong, Y., Choosai, C., Panyim, S., & Damak, S. (2002). Transovarial transmission of sugarcane white leaf phytoplasma in the insect vector *Matsumuratettix hiroglyphicus* (Matsumura). *Insect Molecular Biology*, 11(1), 97–103. <https://doi.org/10.1046/j.0962-1075.2001.00314.x>
- Hanboonsong, Y., Ritthison, W., Choosai, C., & Sirithorn, P. (2006). Transmission of sugarcane white leaf phytoplasma by *Yamatotettix flavovittatus*, a new leafhopper vector. *Journal of Economic Entomology*, 99(5), 1531–1537. <https://doi.org/10.1093/jee/99.5.1531>
- Huang, Z., Bai, X., Gouda, M., Hu, H., Yang, N., & He, Y. (2025). Transfer learning for plant disease detection model based on low-altitude UAV remote sensing. *Precision Agriculture*, 26, 15. <https://doi.org/10.1007/s11119-024-10217-x>
- Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access*, 7, 59069–59080. <https://doi.org/10.1109/ACCESS.2019.2914929>

- Jocher, G., Stoken, A., Borovec, J., NanoCode012, ChristopherSTAN, & Changyu, L. (2021). *ultralytics/yolov5: v6.0 — YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support* (Version v6.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.5563715>
- Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8* (Version 8.0.0) [Computer software]. <https://github.com/ultralytics/ultralytics> (last accessed 8 May 2024).
- Kanapiya, A., Amanbayeva, D., Tulegenova, Z., Abash, A., Zhangazin, S., Dyussebayev, K., et al. (2024). Recent advances and challenges in plant viral diagnostics. *Frontiers in Plant Science*, 15, 1451790. <https://doi.org/10.3389/fpls.2024.1451790>
- Keerthipala, A. P. (2016). Development of sugar industry in Sri Lanka. *Sugar Tech*, 18(6), 612–626. <https://doi.org/10.1007/s12355-016-0485-3>
- Kouadio, L., El Jarroudi, M., Belabess, Z., Laasli, S. E., Roni, M. Z. K., Amine, I. D. I., et al. (2023). A review on UAV-based applications for plant disease detection and monitoring. *Remote Sensing*, 15(17), 4273. <https://doi.org/10.3390/rs15174273>
- Kumarasinghe, N. C., & Jones, P. (2001). Identification of white leaf disease of sugarcane in Sri Lanka. *Sugar Tech*, 3(1–2), 55–58. <https://doi.org/10.1007/BF02945532>
- Lee, I. M., Hammond, R. W., Davis, R. E., & Gundersen, D. E. (1993). Universal amplification and analysis of pathogen 16S rDNA for classification and identification of mycoplasma-like organisms. *Phytopathology*, 83(8), 834–842. <https://doi.org/10.1094/Phyto-83-834>
- Leon-Rueda, W., Ponce, P., Carrera, E. V., & Merizalde, D. (2021). Use of multi-spectral images and vegetation indices to estimate leaf area index and canopy nitrogen content in potato crops in the high Andean zone. In *Proceedings of the 2021 IEEE Fifth Ecuador Technical Chapters Meeting* (pp. 1–6). IEEE, Piscataway, NJ, USA. <https://doi.org/10.1109/ETCM53643.2021.9590662>
- Li, X., Zhu, B., Li, S., Liu, L., Song, K., & Liu, J. (2025). A comprehensive review of crop chlorophyll mapping using remote sensing approaches: Achievements, limitations, and future perspectives. *Sensors*, 25(8), 2345. <https://doi.org/10.3390/s25082345>
- Liu, W., Cao, X., Fan, J., Wang, Z., Yan, Z., Luo, Y., et al. (2018). Detecting wheat powdery mildew and predicting grain yield using unmanned aerial photography. *Plant Disease*, 102(10), 1981–1988. <https://doi.org/10.1094/PDIS-12-17-1893-RE>
- Lu, W. J., Huang, Y. K., Li, W. F., Wang, X. Y., & Luo, Z. M. (2012). Development and application of nested PCR assay for detection of phytoplasma of sugarcane white leaf. *Acta Phytopathologica Sinica*, 42(3), 311–314.
- Mesas-Carrascosa, F. J., Torres-Sánchez, J., Clavero-Rumbao, I., García-Ferrer, A., Peña, J. M., Borra-Serrano, I., et al. (2015). Assessing optimal flight parameters for generating accurate multi-spectral orthomosaics by UAV to support site-specific crop management. *Remote Sensing*, 7(10), 12793–12814. <https://doi.org/10.3390/rs71012793>
- Narmilan, A., Gonzalez, F., Salgado, A. S. A., & Powell, K. (2022). Detection of white leaf disease in sugarcane using machine learning techniques over UAV multi-spectral images. *Drones*, 6(9), 230. <https://doi.org/10.3390/drones6090230>
- Nguyen, C., Sagan, V., Skobalski, J., & Severo, J. I. (2023). Early detection of wheat yellow rust disease and its impact on terminal yield with multi-spectral UAV imagery. *Remote Sensing*, 15(13), 3301. <https://doi.org/10.3390/rs15133301>
- Pacal, I., Kunduracioglu, I., Alma, M. H., Deveci, M., Kadry, S., Nedoma, J., et al. (2024). A systematic review of deep learning techniques for plant diseases. *Artificial Intelligence Review*, 57, 304. <https://doi.org/10.1007/s10462-024-10944-7>
- Padilla, R., Netto, S. L., & da Silva, E. A. B. (2020). A survey on performance metrics for object-detection algorithms. In *Proceedings of the 2020 International Conference on Systems, Signals and Image Processing* (pp. 237–242). IEEE, Piscataway, NJ, USA. <https://doi.org/10.1109/TWSSIP48289.2020.9145130>
- Pix4D SA. (2021). *Pix4Dmapper 4.6 user manual*. Pix4D SA, Lausanne, Switzerland.
- QGIS Development Team. (2021). *QGIS geographic information system* (Version 3.22) [Computer software]. Open Source Geospatial Foundation. <https://qgis.org> (last accessed 10 April 2024).

- Radočaj, D., Radočaj, P., Plaščak, I., & Jurišić, M. (2025). Evolution of deep learning approaches in UAV-based crop leaf disease detection: A Web of Science review. *Applied Sciences*, 15(19), 10778. <https://doi.org/10.3390/app151910778>
- Ram, B. G., Oduor, P., Igathinathane, C., Howatt, K., & Sun, X. (2024). A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, 222, 109037. <https://doi.org/10.1016/j.compag.2024.109037>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788). IEEE, Piscataway, NJ, USA. <https://doi.org/10.1109/CVPR.2016.91>
- Sanseechan, P., Saengprachathanarug, K., Posom, J., Wongpichet, S., Chea, C., & Wongphati, M. (2019). Use of vegetation indices in monitoring sugarcane white leaf disease symptoms in sugarcane field using multispectral UAV aerial imagery. *IOP Conference Series: Earth and Environmental Science*, 301(1), 012025. <https://doi.org/10.1088/1755-1315/301/1/012025>
- Strachan, S., Bhuiyan, S. A., Thompson, N., Nguyen, N. T., Ford, R., & Shiddiky, M. J. A. (2022). Latent potential of current plant diagnostics for detection of sugarcane diseases. *Current Research in Biotechnology*, 4, 475–492. <https://doi.org/10.1016/j.crbiot.2022.10.002>
- Sugar Research Institute, Sri Lanka. (2022). *Annual report 2022*. Sugar Research Institute, Uda Walawe, Sri Lanka.
- Tejaswini, Rastogi, P., Dua, S., Manikanta, & Dagar, V. (2024). Early disease detection in plants using CNN. *Procedia Computer Science*, 235, 3468–3478. <https://doi.org/10.1016/j.procs.2024.04.327>
- Tomaszewski, M., Nalepa, J., Moliszewska, E., Ruszczak, B., & Smykała, K. (2023). Early detection of *Solanum lycopersicum* diseases from temporally-aggregated hyperspectral measurements using machine learning. *Scientific Reports*, 13(1), 7671. <https://doi.org/10.1038/s41598-023-34079-x>
- Tzutalin. (2015). *LabelImg* (Version 1.4.0) [Computer software]. GitHub. <https://github.com/tzutalin/labelImg> (last accessed 15 March 2024).
- Viswanathan, R. (2021). Impact of insect pests and diseases on sugarcane crop production and their management. In R. Viswanathan & C. Chinnaraja (Eds.), *Insect pests and diseases of sugarcane: Diagnostics and management* (pp. 1–28). Scientific Publishers, Jodhpur, India.
- Wang, X. Y., Li, W. F., Huang, Y. K., Zhang, R. Y., Luo, Z. M., Shan, H. L., et al. (2014). Analyses of the 16S-23S intergenic region of the phytoplasma causing the sugarcane white leaf disease in Yunnan Province, China. *Tropical Plant Pathology*, 39(2), 184–188. <https://doi.org/10.1590/S1982-56762014000200010>
- Zeng, W. (2024). Image data augmentation techniques based on deep learning: A survey. *Mathematical Biosciences and Engineering*, 21(6), 6190–6224. <https://doi.org/10.3934/mbe.2024272>
- Zhang, X., Han, L., Dong, Y., Shi, Y., Huang, W., Han, L., et al. (2019). A deep learning-based approach for automated yellow rust disease detection from high-resolution hyperspectral UAV images. *Remote Sensing*, 11(13), 1554. <https://doi.org/10.3390/rs11131554>
- Zhao, X., Zhai, L., Chen, J., Zhou, Y., Gao, J., Xu, W., et al. (2024). Recent advances in microfluidics for the early detection of plant diseases in vegetables, fruits, and grains caused by bacteria, fungi, and viruses. *Journal of Agricultural and Food Chemistry*, 72(28), 15401–15415. <https://doi.org/10.1021/acs.jafc.4c00454>
- Zhu, H., Lin, C., Liu, G., Wang, D., Qin, S., Li, A., et al. (2024). Intelligent agriculture: Deep learning in UAV-based remote sensing imagery for crop diseases and pests detection. *Frontiers in Plant Science*, 15, 1435016. <https://doi.org/10.3389/fpls.2024.1435016>