

Article

Oblique UAV RGB Imagery Improves Rapid Detection of Wilt-Affected Pine Crowns with YOLO11

Yujie Liu ¹, Jinde Ji ¹, Kaihong Xie ¹, Zhongyi Zhan ², Lihua Tao ³, Tingwu Li ¹ and Qi Jiang ^{1,*} 

¹ Pest Research Center, Yinglin Branch, Yunnan Institute of Forest Inventory and Planning, Kunming 650032, China; yujieliu@bjfu.edu.cn (Y.L.)

² College of Forestry, Southwest Forestry University, Kunming 650051, China

³ Yunnan Baiyun Information Technology Co., Ltd., Kunming 650032, China

* Correspondence: qiaoba78@126.com

Abstract

Rapid detection of wilt-affected pine crowns in mountainous forests is hindered by occlusion, self-shadowing, and heterogeneous backgrounds in conventional nadir products. We evaluated whether oblique UAV RGB imagery improves crown-level detection relative to nadir imagery under matched site, season, sensor, and workflow conditions. The workflow was designed for rapid post-flight screening of geotagged UAV photographs. Paired nadir orthophotos and 45–70° oblique photographs were acquired over pine stands in Wenshan Prefecture, Yunnan, China, and organized into D1 (nadir), D2 (oblique), and D3 (simple mixed-view concatenation). Three YOLO11 detectors were trained for crown shoot damage ratio (SDR)-derived operational classes: early-stage ($\text{SDR} < 50\%$), severely damaged ($\text{SDR} \geq 50\%$), and withered (needle-free dead crowns). A paired crown-level RGB subset ($n = 20$ crowns observed in both views) was analyzed as supporting evidence for view-dependent appearance differences. The oblique-image model (D2) achieved the highest validation performance, with precision of 0.994, recall of 0.991, F1-score of 0.989, mAP@0.5 of 0.995, and mAP@0.5:0.95 of 0.880. The paired subset showed a significant multivariate RGB profile difference between views (Hotelling's $T^2 = 58.91$, $F = 3.10$, $p = 0.044$), driven mainly by reduced Excess Green and greater dispersion of blue-related traits under oblique viewing. These results indicate that oblique UAV photographs retain additional crown-edge, lateral-structure, and chromatic context for detecting wilt-affected pine crowns. Oblique RGB imagery therefore provides a practical, low-cost input for rapid forest health surveillance and targeted field verification in rugged pine landscapes.

Keywords: UAV; forest health surveillance; oblique photogrammetry; YOLO11; wilt-affected pine crowns; RGB analysis



Academic Editor: Maciej Pach

Received: 12 April 2026

Revised: 13 May 2026

Accepted: 14 May 2026

Published: 17 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Forest ecosystems are being reshaped by interacting biotic and abiotic disturbances, and the frequency, spatial extent, and ecological severity of tree mortality events are increasing under warming climates, rising vapor pressure deficit, drought intensification, and simplified stand structure [1–3]. In pine-dominated systems, mortality can be driven by pine wilt disease, bark- and wood-boring insects, defoliators, drought, fire, or compound disturbance cascades, yet these processes often converge on a recognizable crown-level syndrome characterized by chlorosis, crown thinning, progressive dehydration, and eventual needle reddening or greying [4–8]. From a surveillance perspective, this convergence

is important: management commonly needs to locate newly symptomatic trees rapidly, even before final attribution to a causal agent is completed in the field.

Rapid detection of wilted, severely damaged, or recently withered pines is operationally valuable for several reasons. First, symptomatic trees may act as local inoculum sources, vector-breeding substrates, or indicators of an actively expanding mortality focus, so delayed recognition can increase the cost and difficulty of sanitation, quarantine, and follow-up investigation [9–12]. This is especially important for trees that have newly entered the withered stage, because they can provide breeding substrates for secondary insects, especially longhorn beetles and bark beetles, and may therefore require prompt sanitation before secondary populations increase. Second, early mapping of newly affected crowns improves prioritization of field verification in landscapes where access is constrained by steep terrain, fragmented roads, and heterogeneous canopy conditions. Third, explicit detection of withered crowns supports post-disaster pest assessment and broader forest-health assessment because the abundance and spatial arrangement of needle-free dead trees provide a directly interpretable indicator of damage severity and recovery priority. Fourth, dry standing dead pines represent a potential fire-hazard component, so locating them can also support fuel-risk screening and removal planning. Symptom mapping can therefore support integrated monitoring systems in which remote sensing, ground diagnosis, hazard reduction, and management response are linked in a time-sensitive workflow rather than treated as independent tasks.

Unmanned aerial vehicles (UAVs) have become a central tool for this type of surveillance because they bridge the gap between labor-intensive field surveys and coarser satellite monitoring [13–15]. UAV platforms can be deployed on demand, can acquire centimetre-scale imagery over difficult terrain, and can produce observations suitable for crown-level analysis. Recent work has shown that UAV-based spectral and image-texture information can support early or intermediate-stage detection of pine decline, particularly when deep learning or advanced machine learning is used to exploit high-resolution crown patterns [16–23]. Broader reviews now emphasize that UAV systems are especially useful where the management objective requires flexible, repeated observations over limited to medium spatial extents and where the target signal is expressed at the individual-tree level [23,24].

Despite these advantages, most operational workflows still depend predominantly on nadir-view orthomosaics. This convention is understandable because nadir imagery is easy to interpret, integrates naturally with mapping workflows, and supports direct generation of orthophotos, canopy height products, or structure-from-motion outputs. However, top-down products also compress three-dimensional crowns into a single viewing geometry. As a result, diagnostically important features such as lateral needle droop, crown-edge collapse, asymmetric thinning, or partially occluded discoloration may be weakened or entirely missed, especially in mountainous forests with strong relief, self-shadowing, and variable background composition [19,25,26]. A nadir orthomosaic can therefore be geometrically convenient while still being suboptimal as a detector input.

This issue has become more important as algorithmic performance has advanced rapidly. Pine wilt detection studies increasingly report strong results from CNNs, semantic segmentation frameworks, and improved YOLO-based detectors trained on UAV imagery [17–22,27,28]. More general reviews of artificial intelligence in forest health surveillance likewise show that disease and pest detection has become one of the dominant application domains for computer vision in forestry [28]. Yet many published gains are difficult to interpret mechanistically because detector architecture, input sensor type, site conditions, and labeling protocols often change simultaneously. Consequently, the specific contribution of image-view geometry remains insufficiently resolved. Put differently, it is

often unclear whether better results arise because the model is stronger, because the sensor is richer, or because the acquisition geometry simply exposes crown symptoms more clearly in the first place.

Oblique UAV imaging is a plausible way to address this limitation. By capturing lateral crown surfaces and additional structural context, oblique photographs can reduce the effective loss of information caused by strict top-down observation. Oblique image networks also align naturally with direct-photo workflows and photogrammetric localization strategies, which may shorten the interval between image acquisition and field response. Previous studies have shown that oblique photogrammetry can improve retrieval of canopy structure proxies such as leaf area index and can enrich structure-from-motion reconstruction by adding complementary viewing angles [25,26]. In the pine wilt domain, the practical value of richer viewing geometry is also reinforced by recent work on unattended UAV hyperspectral systems, multispectral object detection, and explainable or lightweight models, all of which underscore the importance of preserving diagnostically relevant cues while remaining operationally efficient [21,27,29,30].

Against this background, the present study evaluates viewing geometry as the central operational factor while keeping the study area, survey period, UAV platform, SDR-based symptom categories, and training workflow fixed. Specifically, we compare a nadir dataset (D1), an oblique dataset (D2), and a simple mixed-view concatenation dataset (D3) for crown-level detection of wilt-affected pines using YOLO11. We emphasize that D3 is a baseline for naive pooling and not a dedicated multi-view fusion architecture. We also analyze an exploratory paired crown-level RGB subset so that interpretation of detector behavior is anchored in directly observed appearance differences while remaining appropriately cautious about sample size and radiometric effects. Figure 1 summarizes the conceptual contrast between nadir orthophotography and oblique image-based monitoring. Our objectives were to: (i) quantify whether oblique RGB imagery improves rapid crown detection over nadir imagery in the same campaign; (ii) assess whether any advantage is associated with measurable, but exploratory, shifts in crown-level RGB profile; and (iii) evaluate the operational relevance of direct inference on geotagged UAV photographs for rapid forest health surveillance in rugged pine landscapes. We do not attempt causal attribution of wilt symptoms to a single agent, nor do we use the paired RGB subset as proof of a physiological mechanism; instead, it is used as a secondary explanatory complement to the detector comparison.

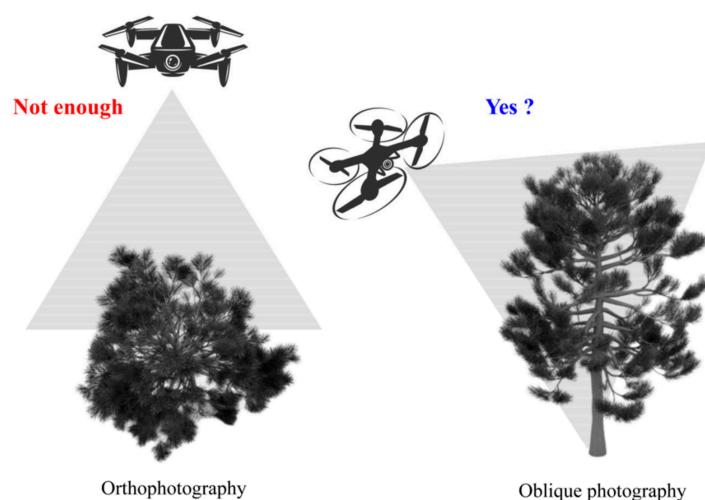


Figure 1. Conceptual comparison of nadir orthophotography and oblique UAV imaging for wilt-affected pine crown detection.

2. Materials and Methods

2.1. Study Site and UAV Data Acquisition

The field campaign was conducted in April 2025 in pine forests distributed across Malipo, Xichou, and Maguan counties, Wenshan Prefecture, Yunnan Province, China. The study region is characterized by mountainous terrain, locally steep slopes, fragmented land cover, and variable crown exposure conditions, all of which create a demanding visual environment for remote detection of individual symptomatic trees. These characteristics make the area an appropriate testbed for evaluating whether image perspective meaningfully alters crown detectability under operational conditions. The study-area context and data-acquisition workflow are summarized in Figure 2.

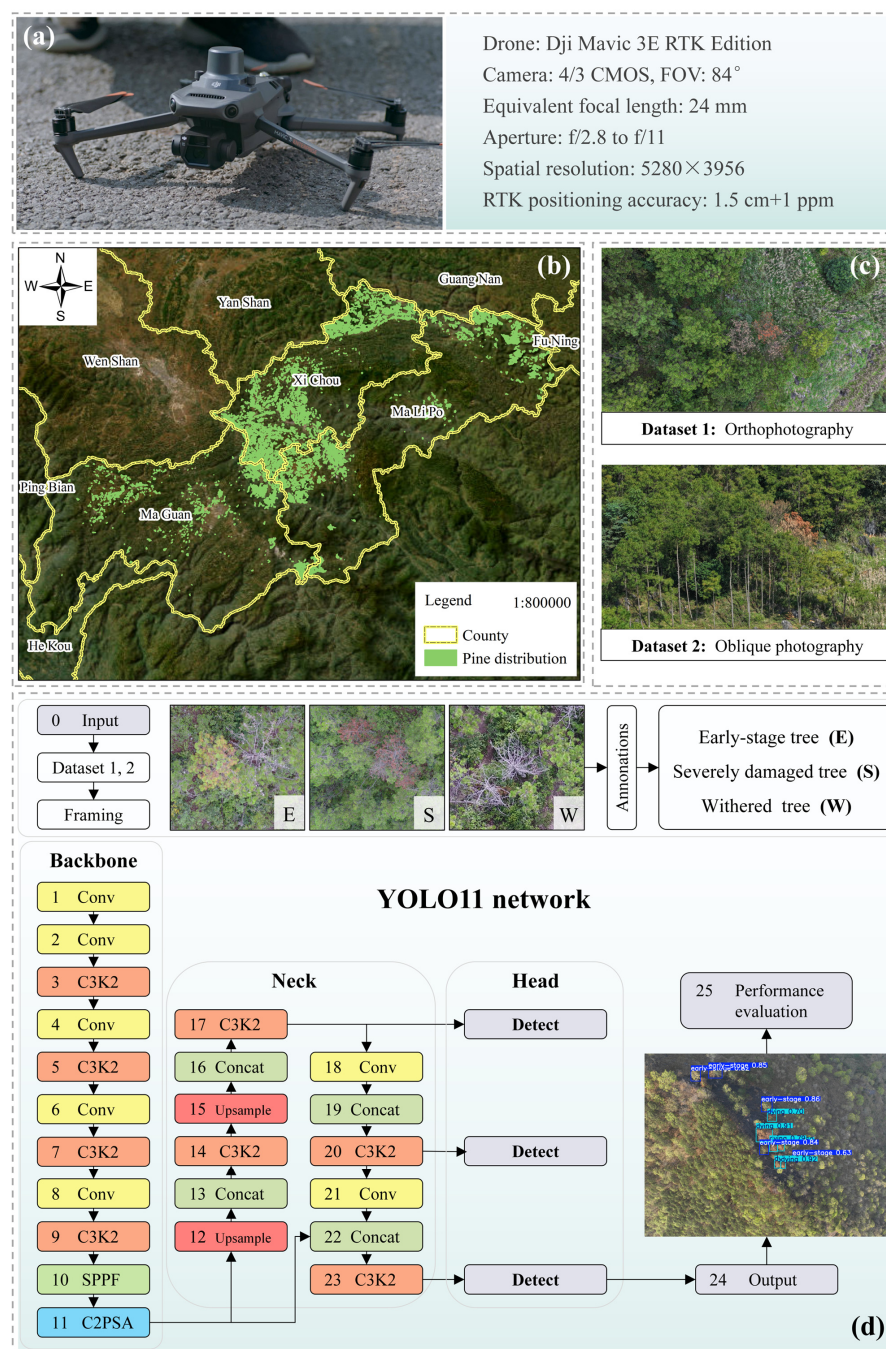


Figure 2. Study overview. (a) UAV platform used for data acquisition (DJI Mavic 3E RTK Edition) and its main camera specifications. (b) Location of the three study counties (Malipo, Xichou, and Maguan)

in Wenshan Prefecture, Yunnan, China, together with the regional pine distribution. (c) Representative examples of the two image datasets acquired over the same pine stands: Dataset 1, nadir orthophotography, and Dataset 2, oblique photography. (d) Overall workflow of the YOLO11-based wilt-pine detection framework, including image acquisition, sample extraction from the two datasets, annotation of three SDR-based symptom-severity classes (early-stage, severely damaged, and withered), model training, and performance evaluation. In the network schematic, C3K2/C3k2 denotes compact CSP-style feature-extraction blocks, and C2PSA denotes the C2-style position-sensitive attention module used in YOLO11. Concat denotes feature concatenation in the revised network schematic.

Images were acquired using a DJI Mavic 3E (RTK edition) equipped with a 4/3 CMOS RGB camera with a 24 mm equivalent focal length (Figure 2a). Terrain-following flight was implemented at an altitude of 150 m, corresponding to an approximate ground sample distance of 4.03 cm. Nadir and oblique images were collected during the same campaign so that viewing geometry could be compared without confounding by season, stand condition, or sensor platform. The oblique dataset was acquired at tilt angles of approximately 45–70°, providing lateral views of crown form in addition to top-visible crown surfaces.

During the same survey period, ground teams geotagged 3217 pine trees and quantified crown damage using the shoot damage ratio (SDR), a widely used metric for crown injury defined in the Standard of Forest Pest Occurrence and Disaster (LY/T 1681-2006) as the proportion of damaged shoots to the total number of shoots in an individual tree crown based on detailed shoot counts [31]. For object detection, the field ratings were consolidated into three operational classes: early-stage (SDR < 50%), severely damaged (SDR ≥ 50%), and withered (trees with no needles). This consolidation was used only to align the detector with operational triage needs; it does not revise the official five-level standard or imply definitive causal diagnosis. Potential decline drivers, including biotic agents, drought, fire, and wind damage, were also recorded to support interpretation of the image-analysis results and crown-level labels in Figure 2.

2.2. Dataset Construction

Two image products were constructed from matched UAV acquisitions over the same stands (Figure 2c). Dataset D1 consisted of nadir-view orthophotography, whereas Dataset D2 consisted of oblique RGB photographs acquired at 45–70° tilt angles. A third dataset, D3, was created by merging D1 and D2 to test whether simple multi-view concatenation improved performance relative to single-view training. The central experimental logic was therefore not to compare different platforms or different sites, but to compare how the same biological target was represented under different viewing geometries.

All datasets were organized for the same three object-detection classes derived from the SDR-based field ratings: early-stage (SDR < 50%), severely damaged (SDR ≥ 50%; abbreviated as ‘severe’ in model outputs), and withered (needle-free crowns). These classes were defined as operational bins along a continuous damage and vigor gradient. Borderline crowns were assigned according to the dominant visible condition, the SDR threshold, and quality-control review. Bounding boxes were assigned at the crown level by trained image analysts using field geotags, SDR-based field records, and visible crown symptoms as anchors. Each annotated image was checked in a second quality-control pass, and ambiguous boxes or severity labels were resolved by consensus with the field records. The validation and test subsets each contained 50 crowns per class, and the training subsets were assembled to preserve an approximately balanced class composition. To reduce leakage from spatial autocorrelation, data partitioning was conducted at the image-block/flight-segment level rather than by randomly splitting individual bounding boxes. Crowns from the same original orthophoto tile, oblique image sequence, or immediately adjacent flight segment were kept in the same subset where possible. Because the imagery came from a single regional campaign with overlapping UAV coverage, residual spatial autocorrelation

cannot be excluded; therefore, the reported metrics are interpreted as within-campaign validation rather than external-site transfer performance.

This dataset design is important for interpretation. If D2 outperformed D1, the advantage could not be attributed to a different site, time period, platform, or symptom taxonomy. However, the comparison should be understood as an operational contrast between nadir and oblique acquisition configurations: oblique imagery changes not only the nominal camera angle but also visible crown sides, background composition, illumination geometry, and self-shadowing. Conversely, if D3 failed to outperform D2, the result would not demonstrate that multiple views are redundant. It would show only that simple sample-level concatenation, without view labels, view-aware embeddings, or explicit cross-view fusion, was insufficient under the present training design.

2.3. YOLO11 Model Development and Evaluation

We adopted the Ultralytics implementation of YOLO11 as a one-stage detector for crown localization and class assignment [29]. YOLO-family detectors are widely used because they combine end-to-end detection with comparatively efficient training and inference, thereby making them attractive for operational workflows and edge-oriented deployment. For clarity, the network diagram in Figure 2d follows the standard YOLO11 notation: C3K2/C3k2 denotes compact CSP-style feature-extraction blocks, whereas C2PSA denotes a C2-style position-sensitive attention module used to strengthen spatially informative features before detection. These modules were used as implemented in YOLO11; no architecture-level innovation was introduced in this study. For context, YOLO11 belongs to a lineage that builds on advances from earlier object detectors including R-CNN, Faster R-CNN, SSD, and the original YOLO framework [32–35]. In this study, the model was not treated as the object of algorithmic innovation; rather, it served as a strong contemporary detector with which to test the effect of image perspective on downstream performance.

Original RGB images (5280×3956 pixels) were resized to 640×640 pixels for training efficiency and to maintain a consistent YOLO input format. This resolution was sufficient for crown-level bounding boxes but may attenuate fine needle-level symptoms; therefore, we interpret the task as crown-scale detection rather than diagnosis of sub-crown physiological detail. Three models were trained separately on D1, D2, and D3 using stochastic gradient descent for 500 epochs with an initial learning rate of 0.01, momentum of 0.937, and weight decay of 0.0005. Non-maximum suppression was used at inference to produce final detections. The same training configuration was retained across datasets so that differences in performance could be attributed as directly as possible to differences in the input data rather than to altered optimization settings. No view-specific hyperparameter tuning or post hoc parameter adjustment was introduced after preliminary comparisons, because doing so would have weakened the fairness of the view-geometry experiment. Inference in this manuscript was performed on post-flight geotagged UAV photographs. We did not conduct an onboard, in-flight latency or frames-per-second benchmark, so the evaluation reports detection accuracy and workflow readiness for rapid screening rather than live autonomous operation.

Model performance was compared using precision, recall, F1-score, mean average precision at an intersection-over-union (IoU) threshold of 0.5 (mAP@0.5), and mean average precision averaged over IoU thresholds from 0.50 to 0.95 in 0.05 increments (mAP@0.5:0.95), together with confusion matrices and loss trajectories. IoU measures the overlap between predicted and reference crown boxes. Precision reflects the fraction of predicted crowns that were correct, recall reflects the fraction of true target crowns detected, and the F1-score summarizes their trade-off. mAP@0.5 reflects average precision under a relatively permissive localization threshold, whereas mAP@0.5:0.95 provides a stricter measure of

localization quality across multiple thresholds. Because the practical goal of this work is timely crown detection, special attention was paid to omission patterns, background confusion, and the stability of validation performance across training, rather than to training loss alone. In deployment, the detector can be applied directly to geotagged UAV photographs, reducing the dependence on full orthomosaic production and thus shortening the lag between data capture and management response.

2.4. Paired Crown-Level RGB Analysis

To provide explanatory support for the detector results, we analyzed the RGB measurements as a paired crown-level dataset. It contained 20 crowns measured in both nadir and oblique views, with one matched observation per view for each crown ID. For each crown, mean red (R), green (G), and blue (B) values and their within-crown standard deviations were available. RSD, GSD, and BSD denote the within-crown standard deviations of red, green, and blue pixel values, respectively. We additionally calculated four RGB-derived indices: Excess Green ($\text{ExG} = 2G - R - B$), Excess Red ($\text{ExR} = 1.4R - G$), $\text{VARI} = (G - R)/(G + R - B)$, and the R/G ratio. These variables were chosen because they summarize different but complementary aspects of crown appearance, including greenness balance, red dominance, and within-crown heterogeneity.

Because the same crowns were observed from both viewpoints, all modality comparisons were treated as paired analyses. For each feature, we calculated the mean paired difference (oblique—nadir), a 95% confidence interval, Wilcoxon signed-rank p value, and paired standardized effect size (d_z). A multivariate paired profile test (Hotelling's T^2) was used to evaluate whether the 10-feature RGB profile differed jointly between viewing modes. Given the modest sample size ($n = 20$ matched crowns) relative to the number of features, this multivariate test was treated as an exploratory profile-level statistic rather than as stand-alone evidence for broad generalization. Principal component analysis (PCA) was then applied to the standardized feature matrix to summarize the dominant axes of crown-level variation. This design is more defensible than an unpaired comparison because it isolates the effect of image geometry from tree-to-tree biological variability. Table 1 presents the main outcomes of this analysis. Because the paired subset comprised only 20 matched crowns, the RGB results were interpreted primarily as explanatory evidence rather than as a stand-alone inferential endpoint.

Table 1. Paired crown-level RGB statistics for nadir and oblique observations ($n = 20$ matched crowns).

Feature	Nadir Mean $\pm$ SD	Oblique Mean $\pm$ SD	Δ (O – N), 95% CI	Wilcoxon p	d_z
R	134.96 $\pm$ 15.69	131.62 $\pm$ 16.87	−3.34 (−11.78, 5.10)	0.330	−0.19
G	124.01 $\pm$ 11.50	121.00 $\pm$ 15.16	−3.01 (−11.28, 5.26)	0.261	−0.17
B	96.45 $\pm$ 7.45	99.54 $\pm$ 19.29	3.09 (−4.68, 10.85)	0.985	0.19
RSD	42.20 $\pm$ 7.12	40.54 $\pm$ 10.03	−1.67 (−5.09, 1.76)	0.409	−0.23
GSD	38.80 $\pm$ 6.98	36.69 $\pm$ 9.27	−2.10 (−5.67, 1.46)	0.294	−0.28
BSD	35.00 $\pm$ 5.93	32.52 $\pm$ 8.20	−2.49 (−5.16, 0.19)	0.070	−0.43
ExG	16.61 $\pm$ 11.89	10.84 $\pm$ 9.24	−5.77 (−9.79, −1.74)	0.011	−0.67
ExR	64.93 $\pm$ 12.22	63.27 $\pm$ 11.41	−1.66 (−6.62, 3.30)	0.349	−0.16
VARI	−0.06 $\pm$ 0.04	−0.07 $\pm$ 0.04	−0.00 (−0.02, 0.02)	0.927	−0.04
R/G	1.09 $\pm$ 0.05	1.09 $\pm$ 0.06	0.00 (−0.03, 0.03)	0.985	0.03

Note: the joint multivariate modality effect across the 10 RGB traits was significant (Hotelling's $T^2 = 58.91$, $F = 3.10$, $p = 0.044$). RSD, GSD, and BSD denote within-crown standard deviations of red, green, and blue pixel values, respectively; ExG, ExR, VARI, and R/G are RGB-derived indices.

The RGB analysis was not intended to replace the detector results, explain neural-network attention with certainty, or claim a complete physiological mechanism of wilt severity. Instead, it was used as a secondary observational bridge between image perspective and detection behaviour. If oblique imagery changed the multivariate crown profile in consistent ways, those shifts could provide a plausible phenotypic context for why one model generalized more cleanly than another within this dataset.

3. Results

3.1. Survey Context

Field observations showed a substantial pool of symptomatic trees across all three severity classes, indicating an actively developing decline landscape rather than a dataset dominated by only terminally dead crowns. The early-stage:severely damaged:withered ratio was approximately 1.15:0.72:1.30, implying that both pre-mortality and post-mortality crowns were well represented during the campaign. This distribution is useful for detector evaluation because it requires the models to handle both subtle and visually conspicuous expression of decline.

When decline or mortality was classified by dominant recorded cause, fire accounted for 27.07% of cases, followed by wind breakage (12.59%) and drought (8.39%). Biotic agents contributed 9.48%, while the remaining observations were attributed to other or uncertain causes. As shown in Figure 3, affected trees were not distributed uniformly but formed localized clusters within the study landscape. These results reinforce that the detector should be interpreted as a symptom-recognition tool rather than a direct etiological classifier. In practice, the workflow is best suited for highlighting crowns that warrant field confirmation, sanitation, or targeted diagnostic follow-up.

3.2. RGB Profile Analysis Based on Paired Crowns

The RGB measurements revealed a modest and exploratory multivariate shift between nadir and oblique observations. When the 10-feature RGB profile (R, G, B, RSD, GSD, BSD, ExG, ExR, VARI, and R/G) was analysed jointly, the paired modality effect was significant (Hotelling's $T^2 = 58.91$, $F = 3.10$, $p = 0.044$). This indicates that the two viewing modes captured measurably different crown appearance profiles in this paired subset, even though the same 20 crowns were analysed in both views. The paired summary statistics are presented in Table 1, and the distributions and multivariate separation are shown in Figure 4. Because the matched sample size was limited, this result should be interpreted as exploratory support for the detector comparison rather than as a definitive generalization about all wilt-affected crowns.

At the single-feature level, however, most channel means remained comparatively stable between modalities. Mean red, green, and blue intensities showed no statistically significant paired differences (Wilcoxon $p = 0.330$, 0.261 , and 0.985 , respectively). Likewise, the R/G ratio and VARI were nearly unchanged on average. The clearest view-dependent shift occurred for Excess Green, which decreased under oblique imaging by 5.77 units on average (95% CI: -9.79 to -1.74 ; Wilcoxon $p = 0.011$; $dz = -0.67$). In practical terms, this suggests that oblique views tended to reduce apparent green dominance without producing a simple, uniform brightness shift across all crowns.

Within-crown heterogeneity showed a more nuanced pattern. Average within-crown standard deviations were slightly lower in oblique images, but the between-crown spread of several traits increased markedly, especially for blue-channel intensity and blue-channel standard deviation. The oblique-to-nadir variance ratio reached 6.70 for mean blue intensity and 1.91 for blue-channel within-crown standard deviation. This means that oblique imagery did not move all crowns in the same direction; rather, it amplified inter-crown

contrast for a subset of blue-sensitive or structurally sensitive traits. Such behaviour is consistent with stronger exposure of shaded crown facets, skeletal branch structure, desiccated tissues, and laterally visible crown sectors that are incompletely captured in top-down views, but it may also partly reflect directional illumination, self-shadowing, and sun-sensor geometry associated with the 45–70° oblique acquisition.

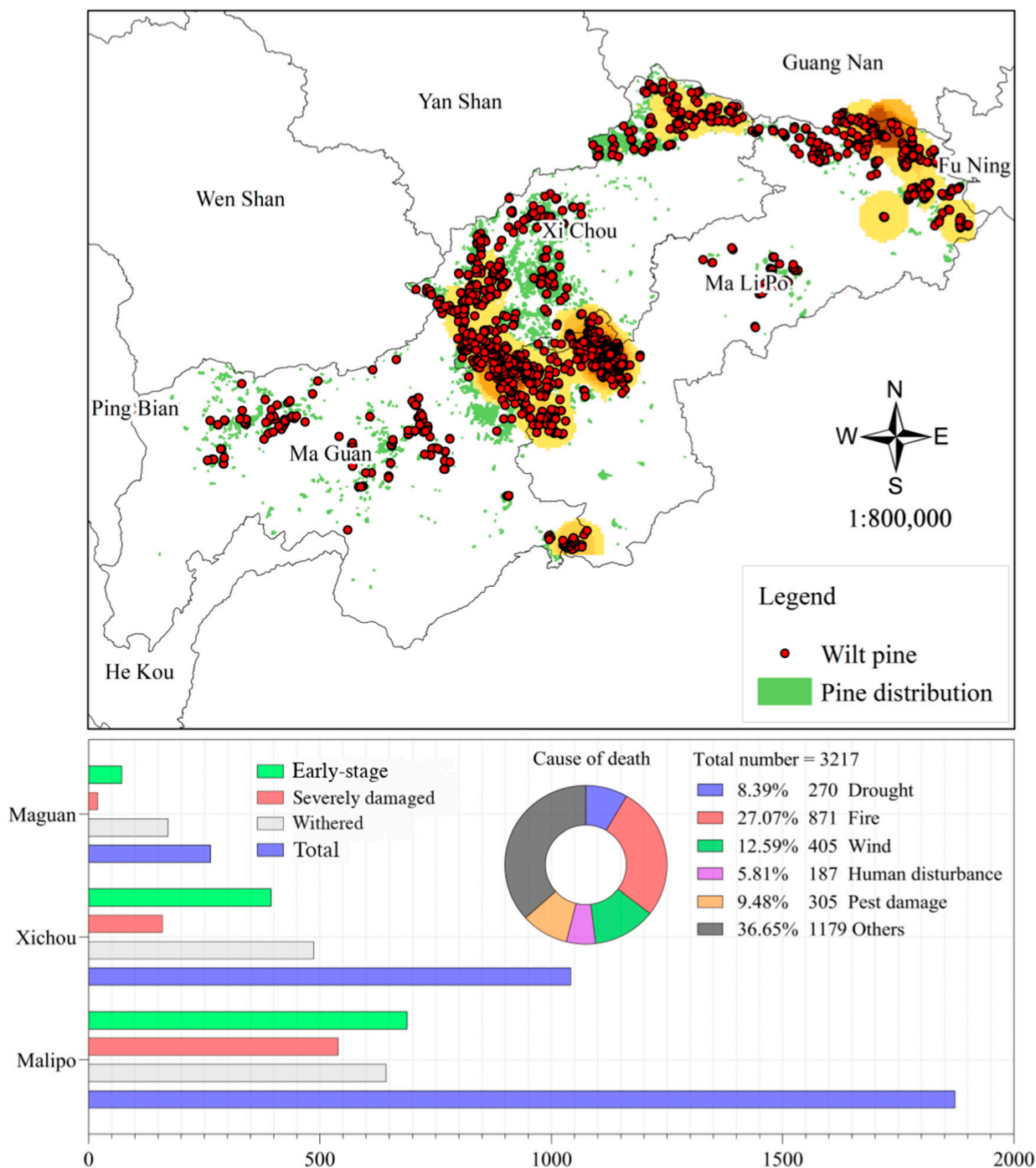


Figure 3. Spatial distribution and field-recorded causes of wilt pine trees in the study area. Note: The orange areas in the map represent the results of kernel density estimation based on the number of wilt pine trees at each sample point, derived using the PLANAR algorithm. A darker color indicates a higher density of wilt pine trees.

PCA further clarified the structure of this shift. As shown in Figure 4D, the first two axes explained 66.7% of the standardized variance (PC1 = 35.7%, PC2 = 30.9%). PC1 was

driven mainly by red-dominance and vegetation-balance variables, especially ExR, R/G, and mean red/green behaviour, whereas PC2 was dominated by the three heterogeneity measures. The partial separation of nadir and oblique observations in PCA space therefore reflects a combined shift in color balance and textural heterogeneity rather than a simple gain in luminance. In other words, oblique images may reweight which parts of crown appearance are most visible to the model, rather than merely changing overall brightness.

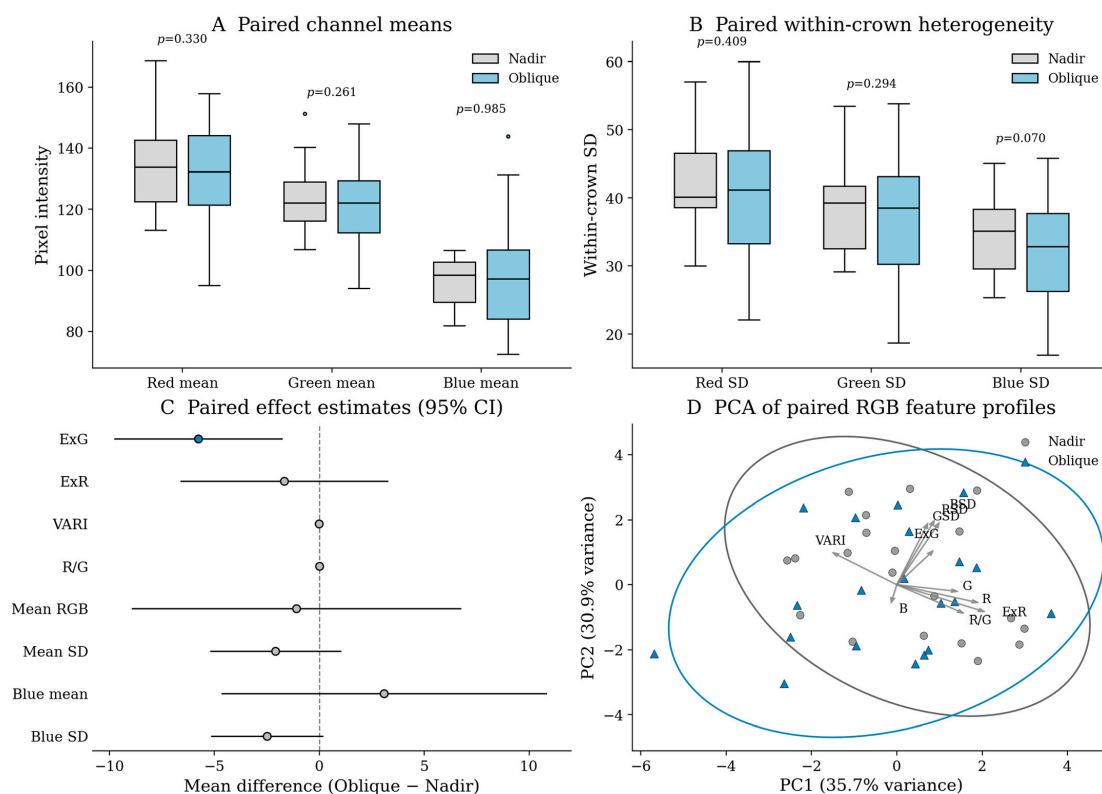


Figure 4. RGB analysis for the paired nadir and oblique crown subset. (A) Paired channel means; (B) paired within-crown heterogeneity; (C) paired effect estimates with 95% confidence intervals; (D) PCA of standardized RGB feature profiles. In panel (C), the blue dot highlights the statistically significant ExG mean paired difference (oblique—nadir), grey dots denote the remaining feature-level estimates, horizontal bars represent 95% confidence intervals, and the vertical dashed line indicates zero difference.

Taken together, the RGB analysis supports a cautious explanatory interpretation: the advantage of oblique imagery did not appear to arise solely because crowns became brighter or more saturated, but because chromatic, shadow-related, and structural cues were re-distributed in a way that plausibly improved crown-level contrast. This interpretation aligns the detector results with directly measured crown-level evidence while avoiding an overextended physiological or causal claim from RGB data alone. The RGB analysis should therefore be read as a parsimonious observational bridge between viewing geometry and downstream detectability, not as proof of how the neural network made every decision.

3.3. Model Performance Comparison

The three YOLO11 models exhibited clear performance differences within the same-campaign validation setting. The oblique-image model (D2) outperformed both the nadir model (D1) and the mixed-view concatenation model (D3) across all major validation metrics (Figure 5A–C). Based on peak validation performance, D2 achieved the highest precision (0.994), recall (0.991), F1-score (0.989), mAP@0.5 (0.995), and mAP@0.5:0.95 (0.880), substantially exceeding the corresponding values of D1 (0.880, 0.799, 0.809, 0.859, and

0.586) and D3 (0.856, 0.821, 0.825, 0.877, and 0.617). The magnitude of the difference in $\text{mAP}@0.5:0.95$ is especially important because it shows that the superiority of D2 extended beyond coarse crown recognition to more precise localization. These high D2 metrics should be interpreted in the context of a controlled, consistently annotated dataset collected under matched site and sensor conditions; they are not an external-transfer benchmark.

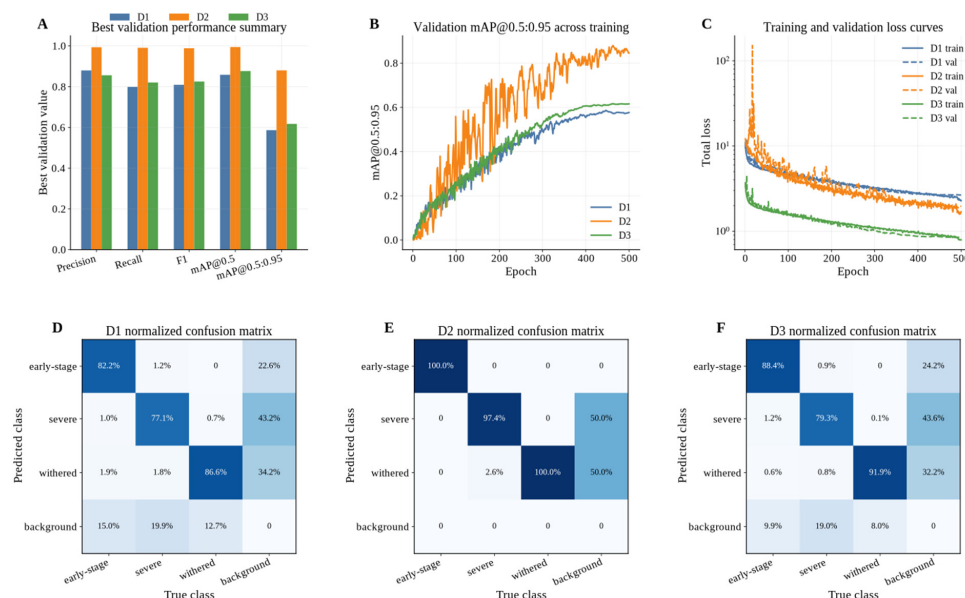


Figure 5. Comparative performance, convergence behavior, and class-level prediction structure of YOLO11 models trained on nadir, oblique, and combined UAV imagery. **(A)** Best validation performance summary, including precision, recall, F1, $\text{mAP}@0.5$, and $\text{mAP}@0.5:0.95$ for the D1, D2, and D3 models. **(B)** Validation $\text{mAP}@0.5:0.95$ trajectories across training epochs for the three models. **(C)** Training and validation loss curves of the D1, D2, and D3 models. **(D)** Normalized confusion matrix of the D1 model trained on nadir imagery. **(E)** Normalized confusion matrix of the D2 model trained on oblique imagery. **(F)** Normalized confusion matrix of the D3 model trained on combined imagery.

The training trajectories in Figure 5B,C reinforced this pattern. D2 rapidly converged toward a substantially higher validation plateau and maintained that advantage through most of the training process. D1 and D3 followed smoother but lower trajectories. Notably, D3, despite containing more total samples, improved only marginally over D1 and remained far below D2. This indicates that under the present non-view-aware training design, simple sample pooling did not yield additive gains. The result should not be interpreted as evidence against multi-view imagery; rather, it suggests that effective multi-view use likely requires explicit view encoding, cross-view correspondence, or dedicated fusion modules.

The confusion matrices in Figure 5D–F explain where these aggregate differences came from. D2 showed near-diagonal dominance, with perfect allocation for the early-stage and withered classes and a severely damaged-class accuracy of 97.4%, with only limited leakage to the withered class. By contrast, D1 and D3 both exhibited substantial target-to-background confusion. In D1, 15.0% of true early-stage crowns, 19.9% of true severely damaged crowns, and 12.7% of true withered crowns were assigned to the background class. D3 reduced some of these errors but still showed considerable omission relative to D2. This means that the performance advantage of D2 was not merely a numerical improvement in a summary statistic; it translated directly into fewer missed symptomatic crowns and cleaner class separation.

From an operational standpoint, this result is important, but its scope should remain bounded. A surveillance system that misses weakly expressed or partially occluded crowns

is less valuable than one that identifies them consistently, even if both systems appear stable during training. Here, oblique imagery yielded the best balance between validation accuracy, omission control, and class separability, indicating that image perspective was a major determinant of detection quality under the conditions of this study. At the same time, the result should be interpreted as evidence for a strong operational advantage in this dataset, not as proof that oblique imagery will always outperform all alternative acquisition designs in other forest types, seasons, or illumination regimes.

3.4. Detection Visualization and Case Analysis

Qualitative inspection of the detection outputs further confirmed the quantitative results. As illustrated in Figure 6, the oblique-trained model detected subtle or partially obscured symptomatic crowns more reliably than the nadir-trained model in several representative situations, including steep slopes, forest edges, occluded crowns, and shaded crowns. The differences were especially evident for weakly expressed early-stage targets and for crowns embedded in visually complex backgrounds. These examples are interpretive rather than statistical, but they are important because they show that the aggregate metric advantage of D2 translated into scenario-level behaviour that is directly relevant for field surveillance.

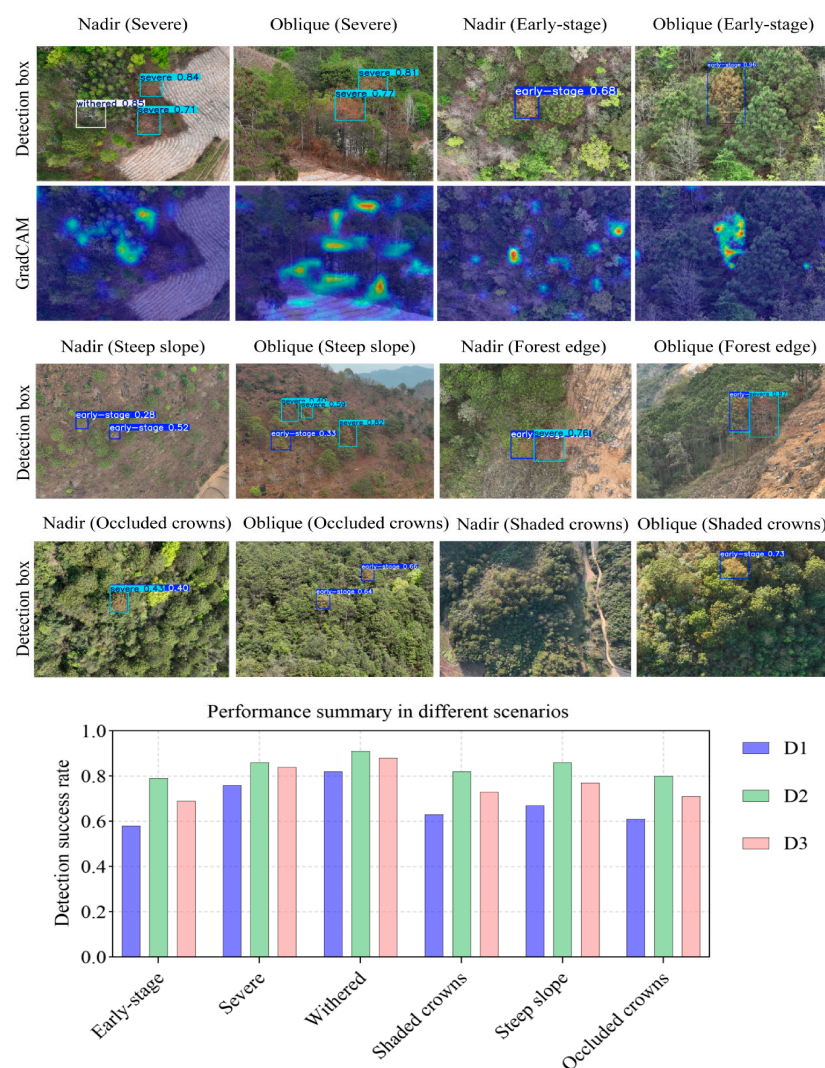


Figure 6. Detection visualizations and scenario-level comparison of the three models. Representative detection outputs are shown for severely damaged and early-stage crowns under nadir and oblique

views, with the corresponding Grad-CAM maps displayed beneath them to illustrate model attention. Additional qualitative detection examples are provided for steep-slope, forest-edge, occluded-crown, and shaded-crown conditions under both viewing geometries. The lower panel summarizes the detection success rates of D1, D2, and D3 across six evaluated categories: early-stage, severely damaged, withered, shaded crowns, steep slope, and occluded crowns.

The Grad-CAM visualizations in Figure 6 provide additional interpretive support. Although Grad-CAM does not furnish a causal explanation and cannot prove which physiological traits the network learned, it can indicate regions of high model sensitivity. D1 tended to emphasize coarse, high-contrast cues such as canopy apices, crown boundaries, and shadow edges, whereas D2 showed more crown-centered responses associated with plausible symptomatic features, including local texture discontinuities, laterally visible crown deformation, and internally heterogeneous crown sectors. The scenario-level summary likewise shows that D2 maintained the highest detection success rate across all tested conditions. Together, these results indicate that oblique imagery improved performance not only in average metrics but also in the specific conditions that matter most for practical forest surveillance.

4. Discussion

4.1. Why Oblique Imagery Improved Detector Performance

The central finding of this study is that operational acquisition geometry was a major determinant of detector performance. Because oblique imagery preserves lateral crown surfaces and retains more three-dimensional symptom expression than nadir products, it likely reduced the compression of diagnostically important features into a top-down canopy silhouette. In wilt-affected crowns, these features include partial discoloration, crown-edge collapse, needle droop, asymmetrical thinning, and localized desiccation. Such cues are often underrepresented in nadir orthomosaics, particularly in mountainous stands where self-shadowing, crown overlap, and background clutter are strong [19,25,26]. However, view geometry should not be read as pure camera angle isolated from all radiometric factors: 45–70° oblique images necessarily change sun-sensor geometry, self-shadowing, visible background, and lateral structural context. The conclusion is therefore that the oblique acquisition configuration outperformed the nadir configuration under matched campaign conditions, not that camera angle alone was mathematically disentangled from illumination and bidirectional reflectance effects. This should not be interpreted as evidence that oblique imagery is universally optimal for every forest-health application; rather, it indicates that when the target symptom is partly expressed on lateral crown surfaces, acquisition geometry can become as consequential as model choice.

This interpretation is consistent with recent UAV-based pine wilt studies showing that detection performance depends strongly on the visual separability of diseased crowns from their surrounding canopy matrix [17–22]. However, most prior studies emphasized model design, sensor selection, or multispectral enhancement. By holding site, campaign period, severity classes, and detector family constant while changing viewing geometry, the present study isolates image perspective itself as an operationally important variable. That distinction matters because it reframes acquisition design as an equal partner to algorithm design. In practical terms, better detector performance may sometimes be achieved more efficiently by improving what the sensor sees than by continuing to increase model complexity alone.

4.2. What the RGB Analysis Adds

The RGB analysis strengthens this study only when interpreted at the appropriate scale. It anchors interpretation in directly measured crown-level observations, but the

paired subset was small and should be treated as exploratory. Importantly, the paired analysis does not support the simplistic idea that oblique imagery merely increases overall brightness or uniformly inflates within-crown variability. Instead, the effect was selective. Green-dominance metrics weakened, blue-related traits became more dispersed across crowns, and the joint RGB profile changed at the multivariate level even when most individual channel means remained stable. Because the matched sample was limited, these patterns should be interpreted as preliminary support for the detector comparison rather than as a complete characterization of crown physiology.

This is a useful explanatory refinement. It suggests that the benefit of oblique imagery may lie less in a global radiometric shift and more in the reweighting of crown appearance cues that help the detector distinguish symptomatic crowns from the surrounding canopy. The stronger spread of blue-related responses may reflect exposure of shaded crown facets, structural skeleton, or desiccated tissues that are incompletely visible from above, while the reduction in ExG is consistent with diminished apparent canopy greenness once lateral and structurally degraded crown sectors are brought into view. Nevertheless, these same patterns may also be influenced by directional shading and canopy reflectance anisotropy. This interpretation remains observational rather than physiological, and it is now deliberately scaled to the available data.

The broader implication is that paired low-cost RGB analysis can provide useful, but preliminary, explanatory support for detector behaviour when more advanced spectroscopy is unavailable. Although RGB data cannot recover the biochemical sensitivity of hyperspectral sensors, it can still reveal whether a different acquisition geometry changes the multivariate crown signature presented to the model. For operational papers, that level of evidence can help move from empirical comparison toward plausible explanation, provided that the limits imposed by sample size, illumination, and radiometric control are stated explicitly.

4.3. Why the Mixed-View Dataset Did Not Outperform the Oblique-Only Model

A noteworthy result is that D3, which combined nadir and oblique samples, remained inferior to D2. This finding should not be interpreted as evidence that multi-view data are useless or biologically redundant. Instead, it shows that the present single-stream YOLO11 training design was not view-aware. From a representation-learning perspective, mixed-view training can enlarge intra-class variance because the same biological class is expressed through substantially different crown geometries, shadow patterns, scale relations, and background structures. Unless the detector is designed to encode view identity, align crowns across views, or fuse features explicitly, that added heterogeneity may dilute the highly informative features already present in oblique imagery.

Operationally, this means that practitioners should not assume that simply collecting more images or pooling more perspectives will necessarily improve performance. The D3 result identifies a limitation of the simplified concatenation strategy rather than a limitation of multi-angle imagery itself. Future work should therefore test dedicated multi-view fusion strategies, such as branch-wise feature extractors, view-aware embeddings, attention-based cross-view aggregation, photogrammetry-guided crown correspondence, or view-consistency losses, rather than relying on dataset concatenation as a substitute for model design.

4.4. Positioning the Findings Within Recent UAV Forest Health Literature

Recent literature provides an informative context for interpreting these results. Reviews of UAV-based forest health monitoring emphasize that UAV systems are particularly powerful when the target signal is expressed at fine spatial scales, when revisit flexibility

is important, and when ground access is difficult [23,24]. Reviews of artificial intelligence in forest health likewise highlight that pest and disease detection has become a leading application domain, but they also note persistent challenges in transferability, explainability, and operational integration [28]. The present study contributes to that discussion by showing that acquisition design itself can be a strong leverage point for improving recognition quality. In that sense, the manuscript sits between sensor-rich early detection studies and low-cost operational RGB workflows: it does not claim that RGB is the most physiologically sensitive modality, but it does show that RGB performance can change materially when view geometry is treated as a designed variable rather than a passive by-product of flight planning.

Within the pine wilt detection literature, multispectral and hyperspectral methods remain essential for early physiological detection and for capturing pre-visual stress signals [16,21,27,30]. In particular, hyperspectral approaches can detect subtle changes before RGB symptoms become visually obvious, and unattended or edge-oriented UAV systems are increasingly being developed to support higher-frequency monitoring [27,30]. Nevertheless, these richer sensors impose higher equipment, calibration, and processing demands. By contrast, RGB sensors remain inexpensive, lightweight, and operationally accessible. The main contribution of this study is therefore not to argue that RGB outperforms hyperspectral data in early physiological diagnosis, but to show that the value of low-cost RGB surveillance can be substantially increased when the imaging geometry is chosen to preserve crown-level structural cues that nadir views tend to suppress.

This distinction is practically important. Many management agencies or field teams can deploy RGB UAVs more readily than hyperspectral systems, especially across rugged, time-sensitive forest landscapes. For those users, the question is often not whether RGB is theoretically the best sensing modality, but how to extract the maximum operational value from RGB imagery that is already available. Our results indicate that, within the tested region and season, oblique acquisition is a practical and immediately actionable way to increase the information content of RGB imagery for crown-level screening.

4.5. Operational Implications and Georeferencing Potential

Beyond detection accuracy, oblique imagery offers a practical route to crown georeferencing and rapid post-flight response. Each oblique photograph is associated with camera position and attitude, and matching a detected crown across overlapping images makes it possible to recover object-space location through standard photogrammetric intersection principles [13,25]. This is particularly attractive for rapid-response workflows because it allows image-level detection and localization to begin before full orthomosaic production is completed for every mission. In mountainous terrain, that reduction in preprocessing overhead can shorten the lag between data capture and field deployment.

In management terms, the proposed workflow is best viewed as a high-throughput triage system rather than an autonomous live-detection system. It can rapidly highlight crowns that warrant field verification, sanitation removal, or causal diagnosis. Importantly, the withered class is not merely a terminal visual category of limited operational value. Trees that have recently reached this stage can become breeding substrates for secondary pests such as longhorn beetles and bark beetles, may require rapid clearing to reduce further pest pressure, provide essential evidence for post-disaster pest assessment and forest-health assessment, and contribute dry standing fuel that can increase local fire risk. Such a role is fully consistent with integrated forest health surveillance and early-warning systems, where UAV detection, field confirmation, sanitation, risk evaluation, and epidemiological interpretation operate as complementary steps rather than as substitutes for one another.

4.6. Limitations and Future Directions

Several limitations should be acknowledged. First, the modelling and image analysis were conducted within one mountainous region and one acquisition season; there was no independent external hold-out site, and geographic transferability across pine species, canopy densities, illumination conditions, seasonal stages, and flight settings remains to be tested explicitly. Second, the training/validation/test partition was designed to reduce spatial leakage at the image-block/flight-segment level, but residual spatial autocorrelation cannot be ruled out in UAV data with overlapping coverage. Third, the detector recognizes symptom expression rather than causal agent identity, which is important in landscapes where drought, fire, wind, and biotic agents can all generate visually similar crown phenotypes. Fourth, the paired RGB data captured only 20 matched crowns and 10 features; Hotelling's T2, PCA, and effect-size results should therefore be considered exploratory, not definitive physiological validation or proof of neural-network attention mechanisms. Fifth, oblique imagery changes not only camera angle but also visible structure, self-shadowing, sun-sensor geometry, background exposure, and bidirectional reflectance effects. Without radiometric calibration, explicit BRDF modelling, or controlled illumination experiments, shading-related contamination of RGB signatures cannot be fully separated from true symptom expression. Sixth, D3 used simple mixed-view concatenation and did not include view labels, view-aware embeddings, cross-view correspondence, or branch-wise fusion. Future work should therefore evaluate external test sites, cross-season campaigns, latency benchmarking on field hardware, radiometrically calibrated imaging, explicit multi-view fusion networks, and uncertainty-aware annotation protocols.

Future work should extend the workflow across seasons, host conditions, and forest types; test external hold-out sites; integrate multispectral, hyperspectral, thermal, or 3D crown attributes; evaluate explicit multi-view fusion networks; and benchmark inference latency on defined field hardware. A particularly promising direction is the combination of oblique RGB detection with lightweight or explainable models that are suitable for onboard, low-latency deployment after speed and power constraints are quantified [21,27,28]. Another is to formalize crown matching across overlapping oblique images so that photogrammetric localization becomes an integral output of the detector rather than a downstream add-on. A further priority is controlled angle optimization, in which 30°, 45°, 60°, and mixed-angle strategies are compared under the same labeling and training protocol. These steps would help determine whether the strong advantage of oblique imagery observed here remains stable across broader operational contexts and whether it can be leveraged for both faster surveillance and more precise intervention.

5. Conclusions

This study demonstrates that oblique UAV RGB imagery provides a substantially stronger basis for rapid detection of wilt-affected pine crowns than conventional nadir imagery within the tested mountainous region and acquisition campaign. Using the SDR-derived early-stage, severely damaged, and withered crown classes, and under matched acquisition conditions, the YOLO11 model trained on oblique images achieved the highest precision, recall, F1-score, mAP@0.5, and mAP@0.5:0.95, and it also showed the cleanest confusion structure and the fewest missed symptomatic crowns. The advantage of oblique imagery therefore reflects not only better aggregate metrics, but also more reliable detection under the visually difficult conditions that matter most in operational forest surveillance. These findings should be viewed as strong within-campaign evidence rather than as an external validation across all seasons, forest types, or illumination conditions.

The exploratory RGB analysis provides a plausible explanation for why this advantage emerged. Oblique imagery altered the joint RGB feature profile of crowns, reduced apparent

green dominance, and expanded the dynamic range of blue-sensitive crown responses, thereby plausibly improving crown-level discriminability without relying on large shifts in average brightness alone. However, these patterns may also reflect directional shading, canopy reflectance anisotropy, and the small paired RGB sample, so they should not be overinterpreted as a definitive physiological mechanism. Taken together, these results support the prioritization of oblique UAV acquisition for rapid forest health surveillance in rugged pine landscapes and show that acquisition geometry should be treated as a core design variable, not merely a by-product of flight planning, when developing practical detector pipelines for wilt-affected forests. The explicit withered class further extends the workflow from early symptom screening to sanitation prioritization, post-disaster and forest-health assessment, and fire-hazard reconnaissance. Even so, broader cross-site validation, latency benchmarking, explicit multi-view fusion, and radiometric calibration remain essential next steps before operational transfer beyond the tested campaign.

Author Contributions: Conceptualization, Y.L. and Q.J.; methodology, Y.L.; field investigation, Y.L., J.J., K.X., and T.L.; formal analysis, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L., Z.Z., L.T., and Q.J.; supervision, Q.J. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by: (1) the Research Project of Yunnan Institute of Forest Inventory and Planning, “Key Technologies for Early Detection of Major Forest Pest Outbreaks in Yunnan Province”; and (2) the Yunnan Provincial Department of Science and Technology Youth Program, “Key Technologies for Early Monitoring and Early-Warning of Major Forest Insect Disasters”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The paired RGB measurements analysed in this study, together with additional model outputs and figure-source files required to reproduce the reported comparisons, can be made available by the corresponding author upon reasonable request.

Acknowledgments: The authors thank the field crews and technical staff who supported UAV acquisition, ground assessment, and dataset preparation during the Wenshan campaign.

Conflicts of Interest: Author Lihua Tao was employed by Yunnan Baiyun Information Technology Co., Ltd.. Author Lihua Tao has received research grants from Yunnan Baiyun Information Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Seidl, R.; Thom, D.; Kautz, M.; Martin-Benito, D.; Peltoniemi, M.; Vacchiano, G.; Wild, J.; Ascoli, D.; Petr, M.; Honkaniemi, J.; et al. Forest disturbances under climate change. *Nat. Clim. Change* **2017**, *7*, 395–402. [[CrossRef](#)] [[PubMed](#)]
2. Pureswaran, D.S.; Roques, A.; Battisti, A. Forest insects and climate change: Physiology, phenology and distribution. *Curr. For. Rep.* **2018**, *4*, 35–50. [[CrossRef](#)]
3. Jactel, H.; Koricheva, J.; Castagneyrol, B. Tree diversity and forest resistance to insect pests: Patterns, mechanisms, and prospects. *Annu. Rev. Entomol.* **2021**, *66*, 277–296. [[CrossRef](#)] [[PubMed](#)]
4. Xu, Q.; Zhang, X.; Li, J.; Ren, J.; Ren, L.; Luo, Y. Pine Wilt Disease in Northeast and Northwest China: A Comprehensive Risk Review. *Forests* **2023**, *14*, 174. [[CrossRef](#)]
5. Zhao, J.; Huang, J.; Yan, J.; Fang, G. Economic Loss of Pine Wood Nematode Disease in Mainland China from 1998 to 2017. *Forests* **2020**, *11*, 1042. [[CrossRef](#)]
6. Lu, J.; Zhao, T.; Ye, H. The shoot-feeding ecology of three *Tomicus* species in Yunnan Province, southwestern China. *J. Insect Sci.* **2014**, *14*, 37. [[CrossRef](#)]
7. McDowell, N.G.; Sapes, G.; Pivovarov, A.; Adams, H.D.; Allen, C.D.; Anderegg, W.R.L.; Arend, M.; Breshears, D.D.; Brodribb, T.; Choat, B.; et al. Mechanisms of woody-plant mortality under rising drought, CO₂ and vapour pressure deficit. *Nat. Rev. Earth Environ.* **2022**, *3*, 294–308. [[CrossRef](#)]

8. Hartmann, H.; Moura, C.F.; Anderegg, W.R.L.; Ruehr, N.K.; Salmon, Y.; Allen, C.D.; Arndt, S.K.; Breshears, D.D.; Davi, H.; Galbraith, D.; et al. Research frontiers for improving our understanding of drought-induced tree and forest mortality. *New Phytol.* **2018**, *218*, 15–28. [\[CrossRef\]](#)
9. Kobayashi, F.; Yamane, A.; Ikeda, T. The Japanese pine sawyer beetle as the vector of pine wilt disease. *Annu. Rev. Entomol.* **1984**, *29*, 115–135. [\[CrossRef\]](#)
10. Togashi, K.; Shigesada, N. Spread of the pinewood nematode vectored by the Japanese pine sawyer: Modeling and analytical approaches. *Popul. Ecol.* **2006**, *48*, 271–283. [\[CrossRef\]](#)
11. Takasu, F. Individual-based modeling of the spread of pine wilt disease: Vector beetle dispersal and the Allee effect. *Popul. Ecol.* **2009**, *51*, 399–409. [\[CrossRef\]](#)
12. Zhou, H.W.; Xie, M.; Koski, T.M.; Li, Y.S.; Zhou, H.; Song, J.Y.; Gong, C.Q.; Fang, G.F.; Sun, J.H. Epidemiological model including spatial connection features improves prediction of the spread of pine wilt disease. *Ecol. Indic.* **2024**, *163*, 112103. [\[CrossRef\]](#)
13. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [\[CrossRef\]](#)
14. Yao, H.; Qin, R.; Chen, X. Unmanned aerial vehicle for remote sensing applications—A review. *Remote Sens.* **2019**, *11*, 1443. [\[CrossRef\]](#)
15. Ecke, S.; Dempewolf, J.; Frey, J.; Schwaller, A.; Endres, E.; Klemmer, K.; Seifert, T. UAV-based forest health monitoring: A systematic review. *Remote Sens.* **2022**, *14*, 3205. [\[CrossRef\]](#)
16. Yu, R.; Luo, Y.; Zhou, Q.; Zhang, X.; Wu, D.; Ren, L. Early detection of pine wilt disease using deep learning algorithms and UAV-based multispectral imagery. *For. Ecol. Manag.* **2021**, *497*, 119493. [\[CrossRef\]](#)
17. Huang, J.; Lu, X.; Chen, L.; Sun, H.; Wang, S.; Fang, G. Accurate identification of pine wood nematode disease with a deep convolution neural network. *Remote Sens.* **2022**, *14*, 913. [\[CrossRef\]](#)
18. Zhou, Y.; Liu, W.; Bi, H.; Chen, R.; Zong, S.; Luo, Y. A detection method for individual infected pine trees with pine wilt disease based on deep learning. *Forests* **2022**, *13*, 1880. [\[CrossRef\]](#)
19. Lee, M.G.; Cho, H.B.; Youm, S.K.; Kim, S.W. Detection of pine wilt disease using time series UAV imagery and deep learning semantic segmentation. *Forests* **2023**, *14*, 1576. [\[CrossRef\]](#)
20. Anwander, J.; Brandmeier, M.; Paczkowski, S.; Neubert, T.; Paczkowska, M. Evaluating Different Deep Learning Approaches for Tree Health Classification Using High-Resolution Multispectral UAV Data in the Black Forest, Harz Region, and Göttinger Forest. *Remote Sens.* **2024**, *16*, 561. [\[CrossRef\]](#)
21. Xu, X.; Zheng, Y.; Gao, Z.; Huang, C.; Wang, X.; Dong, D. Automatic pine wilt disease detection based on improved YOLOv8 using UAV multispectral imagery. *Ecol. Inform.* **2024**, *86*, 102846. [\[CrossRef\]](#)
22. Yang, C.; Lu, J.; Fu, H.; Guo, W.; Shao, Z.; Li, Y.; Zhang, M.; Li, X.; Ma, Y. Detection of pine wilt disease-infected dead trees in complex mountainous areas using enhanced YOLOv5 and UAV remote sensing. *Remote Sens.* **2025**, *17*, 2953. [\[CrossRef\]](#)
23. Manase, A.; Manyevere, A.; Abd Elbasit, M.A.M.; Mashamaite, C.V. The use of UAV-based systems in monitoring forest health: Potentials and challenges. *Sci. Afr.* **2025**, *28*, e02724. [\[CrossRef\]](#)
24. Spiers, A.I.; Scholl, V.M.; McGlinchy, J.; Balch, J.; Cattau, M.E. A review of UAS-based estimation of forest traits and characteristics in landscape ecology. *Landsc. Ecol.* **2025**, *40*, 29. [\[CrossRef\]](#)
25. Lin, L.; Yu, K.; Yao, X.; Deng, Y.; Hao, Z.; Chen, Y.; Wu, N.; Liu, J. UAV-based estimation of forest leaf area index (LAI) through oblique photogrammetry. *Remote Sens.* **2021**, *13*, 803. [\[CrossRef\]](#)
26. Matsuoka, M.; Moriya, H.; Yoshioka, H. Correction of canopy shadow effects on reflectance in an evergreen conifer forest using a 3D point cloud. *Remote Sens.* **2020**, *12*, 2178. [\[CrossRef\]](#)
27. Sui, M.; Wang, X.; Yang, S.; Li, L.; Li, W.; Nie, C.; Huang, H. Diagnosis of pine wilt disease using unattended UAV hyperspectral imager: A comparison of discrete bands and continuous spectrum-based methods. *Ecol. Inform.* **2026**, *93*, 103589. [\[CrossRef\]](#)
28. Amoah-Nuamah, J.; Child, B.; Okyere, E.Y.; Adams, O.; Danquah, J.A. Applications of artificial intelligence in forest health surveillance and management. *Discov. For.* **2025**, *1*, 56. [\[CrossRef\]](#)
29. Ultralytics. YOLO11 Documentation. Available online: <https://docs.ultralytics.com/models/yolo11/> (accessed on 12 April 2026).
30. Liu, W.; Xie, Z.; Du, J.; Li, Y.; Long, Y.; Lan, Y.; Liu, T.; Sun, S.; Zhao, J. Early detection of pine wilt disease based on UAV reconstructed hyperspectral image. *Front. Plant Sci.* **2024**, *15*, 1453761. [\[CrossRef\]](#)
31. LY/T 1681-2006; Standard of Forest Pests Occurrence and Disaster. China Standards Press: Beijing, China, 2006.
32. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [\[CrossRef\]](#)
33. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the 29th International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.

34. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single shot MultiBox detector. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37. [[CrossRef](#)]
35. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.