remote sensing

MDPI

Article

Post-Hail Damage Assessment in Radiata Pine Plantations Using Integrated UAV LiDAR, Multispectral, and Thermal Sensing

Anthony Finn *, Phillip S. M. Skelton and Dianne Patzel

Forestry Centre of Excellence, Adelaide University, Adelaide, SA 5000, Australia
* Correspondence: anthony.finn@adelaide.edu.au; Tel.: +61-(0)404-856-989

Highlights

What are the main findings?

- NDRE was the strongest standalone discriminator of expert-mapped hail-damage severity in *Pinus radiata*, achieving an AUC of 0.954 for severe damage.
- Combining NDRE with NRCT modestly improved spatially cross-validated classification performance, although this improvement was sensitive to spatial partitioning. Adding LiDAR-derived LAI provided no robust improvement.

What are the implications of the main findings?

- UAV-derived multispectral measurements offer an effective means of rapidly assessing and spatially mapping hail-damage severity in radiata pine plantations.
- Thermal sensing may provide complementary information, while LiDAR supports individual-tree detection and crown delineation, enabling tree-level damage mapping against expert-defined severity zones.

Abstract

Hailstorms can cause substantial structural and physiological damage to commercial forest plantations, potentially necessitating premature salvage harvesting and reducing recovery of higher-value timber products. Rapid, spatially explicit methods are therefore needed to quantify post-hail damage at operational scales. This study evaluated a UAV-based workflow integrating multispectral, thermal, and LiDAR observations over a commercial Pinus radiata plantation affected by severe hail near Casterton, Victoria, Australia. The normalised difference red edge index (NDRE) and normalised relative canopy temperature (NRCT) were evaluated alongside LiDAR-derived leaf area index (LAI) and crown diameter. Spectral and structural information was additionally combined into an integrated canopy condition index (ICCI; NDRE × LAI) and an RGB-based variant (ICCIG; GLI × LAI). Expert-delineated polygons representing predominant area-level severity of low, medium, and severe crown damage were used to assign reference classifications to 151,022 individual trees. NDRE was the strongest standalone predictor of polygon-derived damage classes, declining from a mean of 0.651 in the low-damage class to 0.398 in the severe-damage class and achieving an AUC of 0.954 for discriminating trees assigned to expert-defined severe-damage zones. Although NRCT showed relatively weak discrimination alone, its addition to NDRE modestly increased spatially cross-validated balanced accuracy against the polygon-derived reference labels from 0.651 to 0.663; however, this improvement was not maintained under more stringent directional spatial validation. Adding LiDAR-derived LAI and crown diameter provided little additional discriminatory information, although LiDAR supported individual-tree detection, crown delineation, and inventory attribution. These findings demonstrate that multispectral red-edge information provides the strongest basis for reproducing expert-mapped post-hail damage severity, while thermal and LiDAR

Check for updates

Academic Editor: Sheng Nie

Received: 16 October 2026
Revised: 20 November 2026
Accepted: 25 November 2026
Published: 28 November 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) license.

observations offer complementary capabilities for operational plantation assessment. The reported classification performance reflects agreement with expert-defined damage zones rather than independently validated individual-tree damage severity.

Keywords: hail damage; UAV remote sensing; airborne LiDAR; multispectral imagery; thermal imaging; NDRE; NRCT; integrated canopy condition index (ICCI); individual-tree classification; plantation forestry

1. Introduction

Hailstorms are among the most damaging weather events affecting agricultural and forest ecosystems, causing substantial economic losses worldwide [1]. However, despite their recognised impacts, quantitative assessment of hail damage in commercial forestry plantations remains comparatively limited.

In Australia, severe hail events occur periodically but unpredictably, including within the Green Triangle and other major plantation regions, where they can cause substantial losses through defoliation, crown breakage, bark injury, and stem damage, with subsequent reductions in timber quality and growth. Historical studies have shown that severe hail can result in extensive salvage harvesting and long-term reductions in plantation productivity, while injuries to stems and branches can increase susceptibility to fungal infection by *Diplodia sapinea* and other secondary stresses [2,3]. Despite these potentially significant economic consequences, post-hail assessment in commercial forests continues to rely largely on visual field inspection and expert judgement, providing limited quantitative information on the spatial distribution, extent, and severity of damage across affected stands.

Conventional post-disturbance assessment relies predominantly on field inspection and expert judgement to inform salvage planning, insurance assessment, and other operational decisions. Although field surveys provide essential reference information, plot-based assessments sample only a proportion of the affected stand and may be difficult to implement rapidly when damage is spatially dispersed across the landscape. Access to damaged forests may also be restricted immediately following severe weather, further delaying assessment when timely information is required for operational decision making. Visual assessment of crown condition can additionally be influenced by the location and nature of damage within the crown and by differences among observers, introducing uncertainty into comparisons among surveys and through time [4]. Consequently, there is a need for rapid, objective, and repeatable methods capable of quantifying both the severity and spatial extent of hail damage across plantation areas.

Recent advances in unmanned aerial vehicle (UAV) technology enable the acquisition of multispectral imagery, thermal observations, and aerial LiDAR, each modality providing complementary information on vegetation physiology and canopy structure [5]. Integrating these datasets offers considerable potential for rapid, objective assessment of post-hail damage.

Despite these advances, relatively little research has investigated the combined use of structural and spectral information for quantitative assessment of hail damage in commercial plantations. Existing studies typically focus either on vegetation indices that characterise physiological condition or on LiDAR-derived structural attributes that quantify physical damage. Integrating these complementary observations offers the potential to characterise both mechanical canopy damage and physiological stress within a common assessment framework.

This study forms part of a post-hail assessment of two commercial plantations—a *Pinus radiata* plantation and a *Eucalyptus globulus* plantation—in the Green Triangle region

of south-eastern Australia, for which UAV-acquired multispectral, thermal, and structural observations were collected following a severe hailstorm. The present paper focuses specifically on the *P. radiata* plantation, while analysis of the *Eucalyptus* plantation will be reported separately. Spectral vegetation indices, relative canopy temperature, and LiDAR-derived structural metrics are evaluated at the individual-tree scale and integrated to quantify the spatial extent and severity of hail damage. Expert-derived mapping of crown-damage severity is used as the operational reference for calibrating empirical damage thresholds and evaluating the ability of spectral, structural, and integrated canopy metrics to reproduce low, medium, and severe crown-damage classes at the individual-tree scale. The objective of this paper is to develop and evaluate a UAV-based workflow for rapid post-hail damage assessment.

2. Related Work

Although hail has long been recognised as a significant disturbance affecting plantation forestry, relatively few studies have quantified its effects using modern remote sensing techniques. Early Australian investigations by Cremer [2,3] documented the physical and operational consequences of hail injury in Pinus radiata, including the possibility of approximately two damaging hailstorms during a typical 35-year rotation. Minko [6] subsequently demonstrated that hail damage may affect plantation condition for many years through reduced growth, timber degradation, and increased susceptibility to secondary stresses. Hail injury varies with storm intensity, wind direction, tree age, and crown form, and may include needle loss, branch breakage, bark injury, and exposure of underlying tissues. Severe injury can predispose trees to infection by *Diplodia sapinea*, potentially resulting in shoot dieback, reduced growth, and mortality [6,7]. Recent climatological analyses further highlight the heterogeneous and episodic nature of Australian hail events, with hail-prone environments increasing in some regions but declining in others [8]. Despite this established understanding of hail impacts, quantitative assessment of damage severity and spatial extent in commercial plantations remains comparatively underdeveloped.

Advances in UAV remote sensing enable centimetre-scale observations of individual tree crowns using multispectral, thermal, and LiDAR sensors. These capabilities provide opportunities to assess forest condition at spatial resolutions directly relevant to operational management. Fraser and Congalton [9] demonstrated the value of UAV multispectral imagery for fine-scale forest health assessment, while subsequent studies have incorporated automated crown delineation and tree-level health classification [10]. Unlike conventional plot-based surveys, UAV observations can provide spatially continuous information on damage distribution across affected plantation compartments.

Spectral vegetation indices provide indicators of canopy condition through changes in pigment absorption, foliage density, and near-infrared reflectance [11,12]. The normalised difference vegetation index (NDVI) is widely used to characterise vegetation vigour [13], although saturation in dense canopies has encouraged the use of alternative indices, including the Enhanced Vegetation Index (EVI) and red-edge formulations [14]. The normalised difference red edge index (NDRE) is particularly relevant to radiata pine because red-edge reflectance is sensitive to chlorophyll-related variation while retaining sensitivity under relatively dense canopy conditions [15]. Hail-induced needle loss and canopy deterioration may therefore alter NDRE sufficiently to distinguish different levels of crown damage. The Green Leaf Index (GLI), derived from conventional RGB imagery or colourised LiDAR data, provides a simpler alternative where multispectral observations are unavailable. Comparing GLI with NDRE allows the potential operational trade-off between sensor complexity and damage-classification performance to be evaluated.

Thermal observations provide complementary information about canopy condition through their relationship with plant water status, transpiration, and surface energy balance. Reduced stomatal conductance, hydraulic function, or transpiring leaf area can decrease evaporative cooling and increase canopy temperature [16,17]. However, thermal responses are also influenced by radiation, air temperature, humidity, wind, and acquisition time. Relative temperature normalisation can facilitate within-survey comparisons, although elevated temperatures are not specific to hail damage and may reflect other physiological or environmental stresses [18–20]. Thermal imagery therefore offers a potentially useful, but environmentally sensitive, indicator of post-disturbance canopy condition.

Individual-tree assessment is particularly relevant because hail damage can vary substantially among neighbouring trees owing to differences in storm exposure, crown form, and local conditions [21]. UAV laser scanning (ULS) provides three-dimensional measurements of crown structure, enabling attributes such as tree height, crown diameter, crown volume, and foliage distribution to be quantified [22–25]. Automated point-cloud analysis further supports individual-tree inventories and structural characterisation across operational forest areas [26,27]. Leaf area index (LAI), canopy cover (CC) and gap fraction (GF) provide related measures of canopy density and openness that can be estimated using LiDAR penetration metrics and gap-fraction theory [28,29]. Hail-induced needle and branch loss may reduce LAI and CC while increasing GF [28–30], potentially providing structural evidence of disturbance that complements spectral observations of the remaining foliage.

Hail damage involves both mechanical injury and subsequent physiological responses, including changes in photosynthetic activity and plant water relations [31]. Consequently, multispectral imagery and LiDAR characterise different manifestations of the same disturbance: spectral measurements provide information about canopy condition, whereas LiDAR quantifies crown architecture and foliage distribution. Previous studies have demonstrated the potential benefits of integrating structural and spectral observations for forest disturbance and health assessment [32,33]. Nevertheless, whether such integration improves discrimination of hail damage severity beyond individual spectral or structural indicators remains an empirical question, particularly in commercial radiata pine plantations.

Composite vegetation indices have been developed to combine complementary spectral information. For example, the Canopy Chlorophyll Content Index (CCCI) [34,35] expresses NDRE relative to NDVI and has been used to characterise canopy chlorophyll-related variation while accounting, in part, for differences in canopy cover. The integrated canopy condition index (ICCI) proposed in this study is conceptually related to the broader objective of combining vegetation-condition indicators but is mathematically distinct from CCCI. Rather than normalising one spectral index by another, ICCI multiplies NDRE by LiDAR-derived LAI to examine whether incorporating canopy-density information improves discrimination of hail damage severity. A second proposed formulation, $ICCI_G$ ($ICCI_G$ = GLI × LAI), substitutes RGB-derived GLI for NDRE, providing an alternative where red-edge imagery is unavailable. Comparing these multiplicative indices with their constituent spectral measurements and additive spectral–structural models enables the incremental contribution of LAI to damage classification to be evaluated directly.

Despite advances in UAV-based forest monitoring, there remains no widely established approach for combining multispectral, thermal, and LiDAR observations with expert-derived reference information to produce empirically calibrated, individual-tree assessments of hail damage in commercial plantations. This study addresses that gap by evaluating spectral, thermal, structural, and integrated canopy metrics against expert-assigned damage classes, deriving quantitative classification thresholds and mapping the severity and spatial extent of post-hail damage.

3. Materials and Methods

3.1. Study Site

Field and UAV surveys were conducted in two commercial plantations near Casterton in the Green Triangle region of south-eastern Australia: a *Pinus radiata* plantation and a *Eucalyptus* plantation (Figure 1). The Green Triangle supports extensive plantations of radiata pine (*Pinus radiata*) and eucalyptus (*Eucalyptus* spp.), managed primarily for timber production. The study sites were selected because they experienced significant damage during a severe hailstorm. The present analysis is restricted to the *P. radiata* plantation; the eucalyptus observations are reserved for separate analysis.

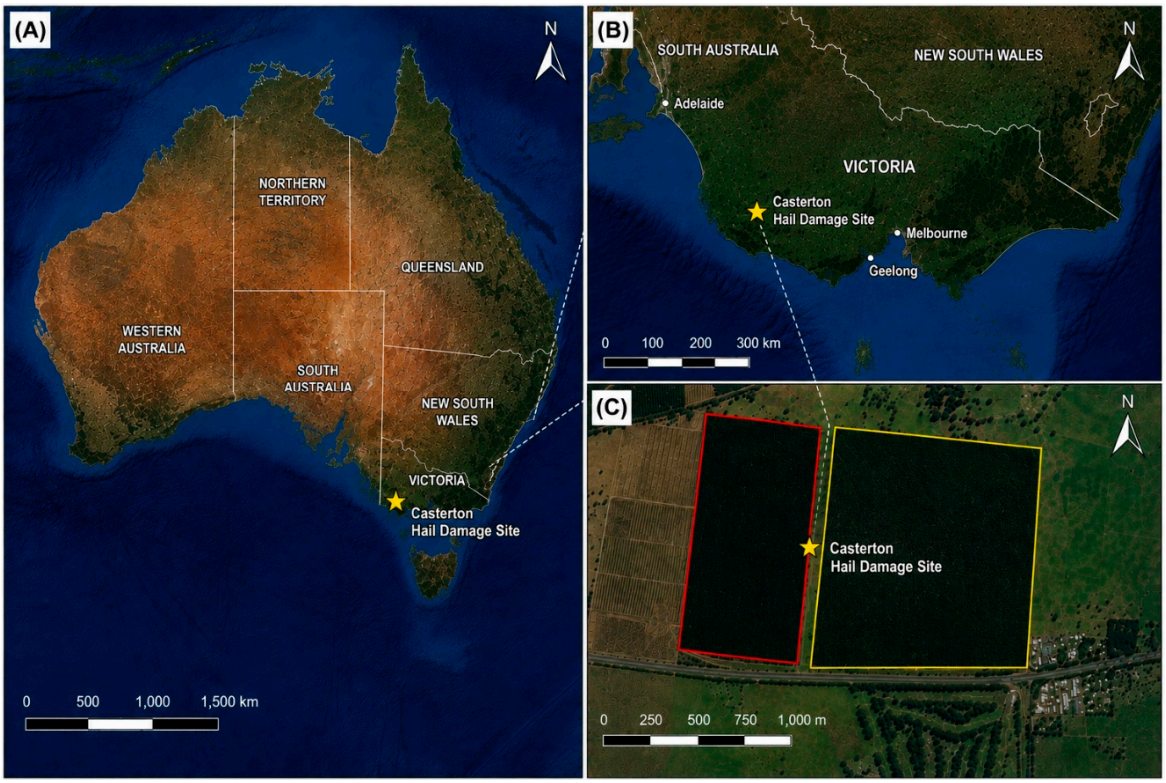

Figure 1. Study area near Casterton, Victoria (VIC), Australia, affected by the October 2024 hailstorm. (**A**) National location within Australia. (**B**) Regional setting in south-western Victoria. (**C**) High-resolution aerial image of the study area showing the eucalyptus plantation (red outline) and adjacent radiata pine plantation (yellow outline) surveyed using UAV multispectral/thermal, RGB and LiDAR sensors.

On 16 October 2024, an intense hailstorm passed through the region impacting approximately 2000 hectares of pine. The storm inflicted substantial mechanical damage across the affected forest. Field observations indicated widespread defoliation and injury to crowns, branches, and stems (Figure 2), with damage severity varying considerably within and between the two sites. This spatial heterogeneity, ranging from relatively lightly affected forest to areas of severe damage, provided a suitable setting for evaluating remotely sensed indicators of hail-damage severity.

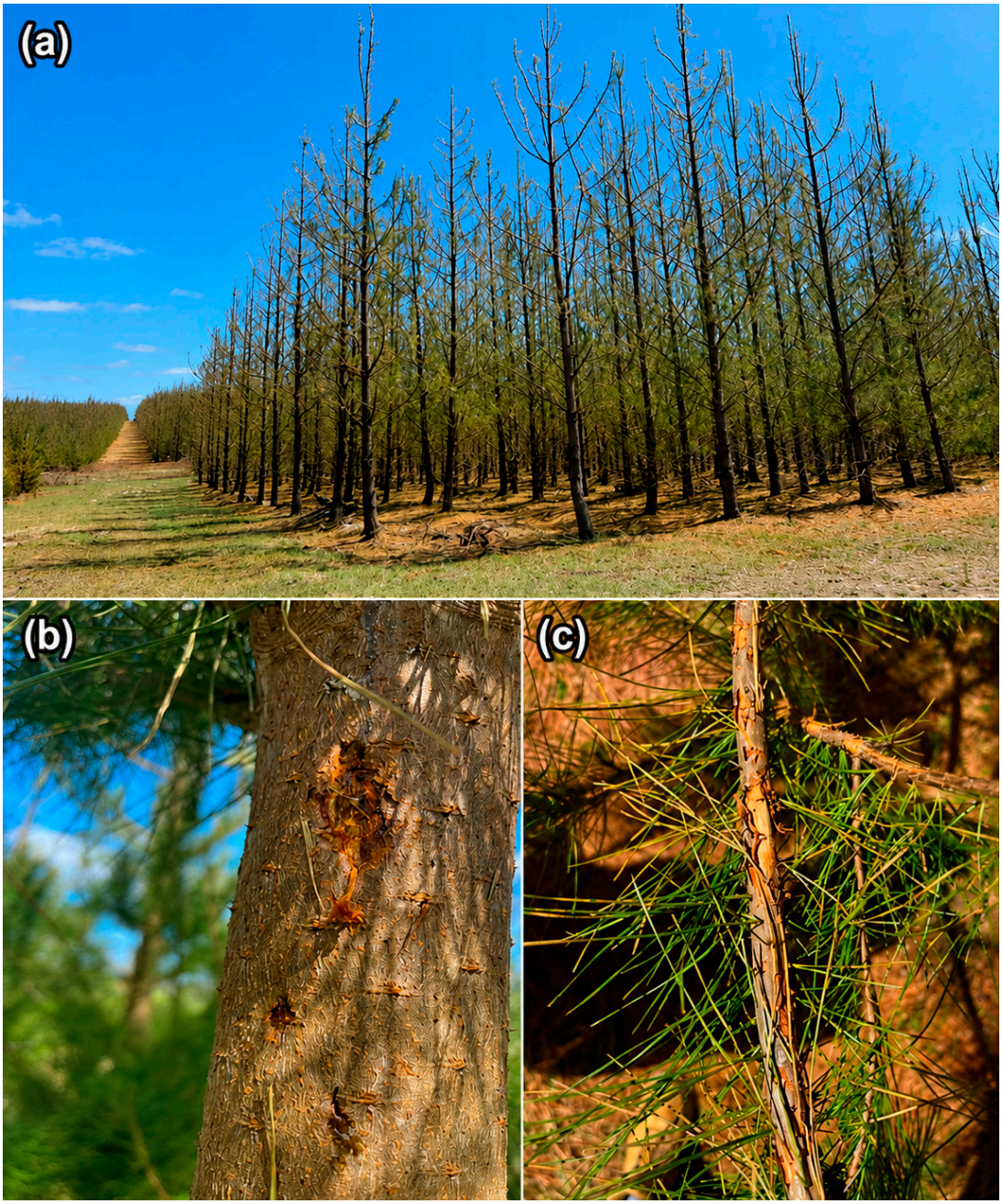

Figure 2. Field examples of hail damage observed following the 16 October 2024 storm near Casterton, Victoria: (**a**) extensive crown defoliation and canopy damage within an affected Pinus radiata stand; (**b**) mechanical injury to the stem, including bark damage and exposure of underlying tissue; and (**c**) hail-induced injury to a branch/shoot with associated foliage damage. Photographs illustrate the range of structural damage observed during post-event field assessment.

3.2. Data Acquisition

Three post-hail UAV surveys were conducted on 8 November 2024, 20 January 2025, and 12 November 2025, accompanied by ground observations of tree condition. The present analysis focuses on the first post-event survey, acquired approximately three weeks after the hailstorm, when structural damage and physiological stress were evident. The subsequent surveys are not analysed here and will be used separately to examine temporal changes in canopy condition and post-hail recovery.

All aerial data were acquired using a DJI Matrice 350 UAV (SZ DJI Technology Co., Ltd., Shenzhen, China) equipped with three complementary sensors: a Zenmuse P1 RGB camera

(www.dji.com), a MicaSense Altum-PT multispectral and thermal camera (www.micasense.com), and a Zenmuse L2 ULS. This combination enabled the simultaneous acquisition of high-resolution optical imagery, multispectral reflectance, thermal observations, and three-dimensional point cloud data for each plantation.

Meteorological conditions on 8 November 2024 were characterised using historical observations from the Australian Bureau of Meteorology (BOM) Casterton weather station (ID 090182). The recorded minimum and maximum air temperatures were 6.2 °C and 16.7 °C, respectively. At 09:00, air temperature was 13.0 °C, relative humidity was 84%, and north-westerly winds were recorded at 19 km/h. By 15:00, air temperature had increased to 15.6 °C, relative humidity had decreased to 56%, and winds were west-south-westerly at 28 km/h. A maximum wind gust of 48 km/h was recorded at 10:04. Rainfall of 4.8 mm was recorded during the 24 h ending at 09:00 on 8 November, with no rainfall recorded during the subsequent 24 h period. These observations provide regional meteorological context for the UAV survey but do not represent direct measurements of atmospheric conditions at the plantation during individual flights.

3.2.1. RGB and Multispectral Imagery

Visible imagery was acquired using the Zenmuse P1 camera (SZ DJI Technology Co., Ltd., Shenzhen, China), while multispectral and thermal data were collected using the MicaSense Altum-PT sensor (MicaSense, Inc., Seattle, WA, USA). The Altum-PT records five multispectral reflectance bands (blue, green, red, red-edge, and near infrared), together with a high-resolution panchromatic channel and long-wave infrared (LWIR) thermal imagery. Although the P1 imagery was acquired as part of the broader survey, and its RGB-derived orthomosaic was used by the forestry assessor to delineate the expert-reference damage polygons (Section 3.3.6), it was not used to derive spectral, thermal, or structural predictor variables evaluated in the classification analyses.

Flights were conducted at an altitude of approximately 100 m above ground level, with 85% forward and side overlap and a flight speed of approximately 7 m s^{-1}, producing ground sampling distances of approximately 1 cm for the RGB imagery and 2.5 cm for the multispectral imagery. Images were processed using Agisoft MetaShape Professional (www.agisoft.com) to generate orthomosaics and georeferenced 3D point clouds using structure-from-motion photogrammetry (SfM) [36].

Radiometric calibration of the MicaSense multispectral imagery was performed using the manufacturer's calibration equipment and procedures. Images of the standard MicaSense calibrated reflectance panel were acquired immediately before each UAV flight to provide a reference for the conversion of recorded image measurements to reflectance. No additional reference panel images were acquired after the flights. All multispectral acquisitions were conducted between approximately 10:00 and 14:00 local time to minimise the effects of low solar elevation angles and associated shadowing. The imagery was processed in Agisoft Metashape using the *Calibrate Reflectance* function, with both the calibrated reflectance panel and the Altum-PT downwelling light sensor (DLS) measurements incorporated into the radiometric calibration. Calibration was applied to the individual images before orthomosaic generation, and the resulting reflectance products were used to calculate spectral vegetation indices, including NDRE, from the near-infrared and red-edge bands. This procedure was intended to reduce radiometric variability arising from differences in sensor response and incident illumination, thereby improving the consistency of spectral measurements across individual tree crowns. Nevertheless, residual effects associated with canopy shadowing, viewing geometry, and local illumination conditions may remain.

3.2.2. LiDAR Data

Three-dimensional structural information was collected using the Zenmuse L2 ULS (SZ DJI Technology Co., Ltd., Shenzhen, China) operating in five-return mode with a pulse repetition frequency of 240 kHz. The manufacturer specifies horizontal and vertical accuracies of about 5 cm and 4 cm, respectively, for a flying height of 150 m. The integrated RGB camera simultaneously recorded colour information for each LiDAR return, allowing colourised point clouds to be generated.

LiDAR surveys were conducted at an altitude of approximately 60 m above ground level using a conventional lawnmower flight pattern, with approximately 50% swath overlap and a flight speed of 10 m s^{-1}. UAV positioning was provided by the Matrice 350 real-time kinematic (RTK) Global Navigation Satellite System (GNSS) system, with real-time differential corrections supplied via the Networked Transport of RTCM via Internet Protocol (NTRIP). The RTK-corrected positioning solution provided centimetre-level georeferencing of the sensor trajectory, with an expected absolute positional accuracy of better than approximately 0.10 m. Raw LiDAR observations were processed using DJI Terra (V5.3.5) to generate georeferenced point clouds, with resulting point densities of approximately 2300 points m^{-2}.

3.3. Data Processing and Analysis

Figure 3 summarises the workflow developed for quantitative post-hail damage assessment. The methodology integrates the UAV-derived multispectral and thermal imagery and ULS data to characterise both the physiological condition and structural response of individual trees following the hail event. LiDAR data were first processed to derive canopy structure, individual-tree crowns, and structural attributes, while multispectral and LWIR observations were independently processed to derive spectral and thermal indicators of vegetation condition. The datasets were subsequently co-registered on a common spatial grid and integrated through the integrated canopy condition index (ICCI, see Section 3.3.5). Expert-assessed damage zones were then used as reference data to empirically calibrate tree-level damage classes (see Section 3.3.6).

3.3.1. LiDAR Preprocessing and Canopy Modelling

The UAV LiDAR dataset comprised about 15 spatially adjacent LAS tiles per site. To minimise memory requirements while avoiding discontinuities at tile boundaries, tiles were processed individually together with neighbouring point-cloud data occurring within a 5 m buffer. Ground returns were identified using LAS ground classifications where available; otherwise, ground points were estimated using spatial down-sampling and the Simple Morphological Filter (SMRF) [37]. A ground surface was interpolated from these returns, and point elevations were normalised to height above ground.

A canopy height model (CHM) was generated at 0.10 m spatial resolution and smoothed prior to tree detection. Tree apices were identified using variable-window local maxima, and individual crowns were subsequently delineated using the watershed segmentation described in Section 3.3.2. Crown polygons were subsequently constructed for each segmented tree. To minimise duplication arising from buffered tile processing, only trees with apices located within the core extent of each tile were retained, after which any residual duplicate apices separated by less than 0.75 m were removed.

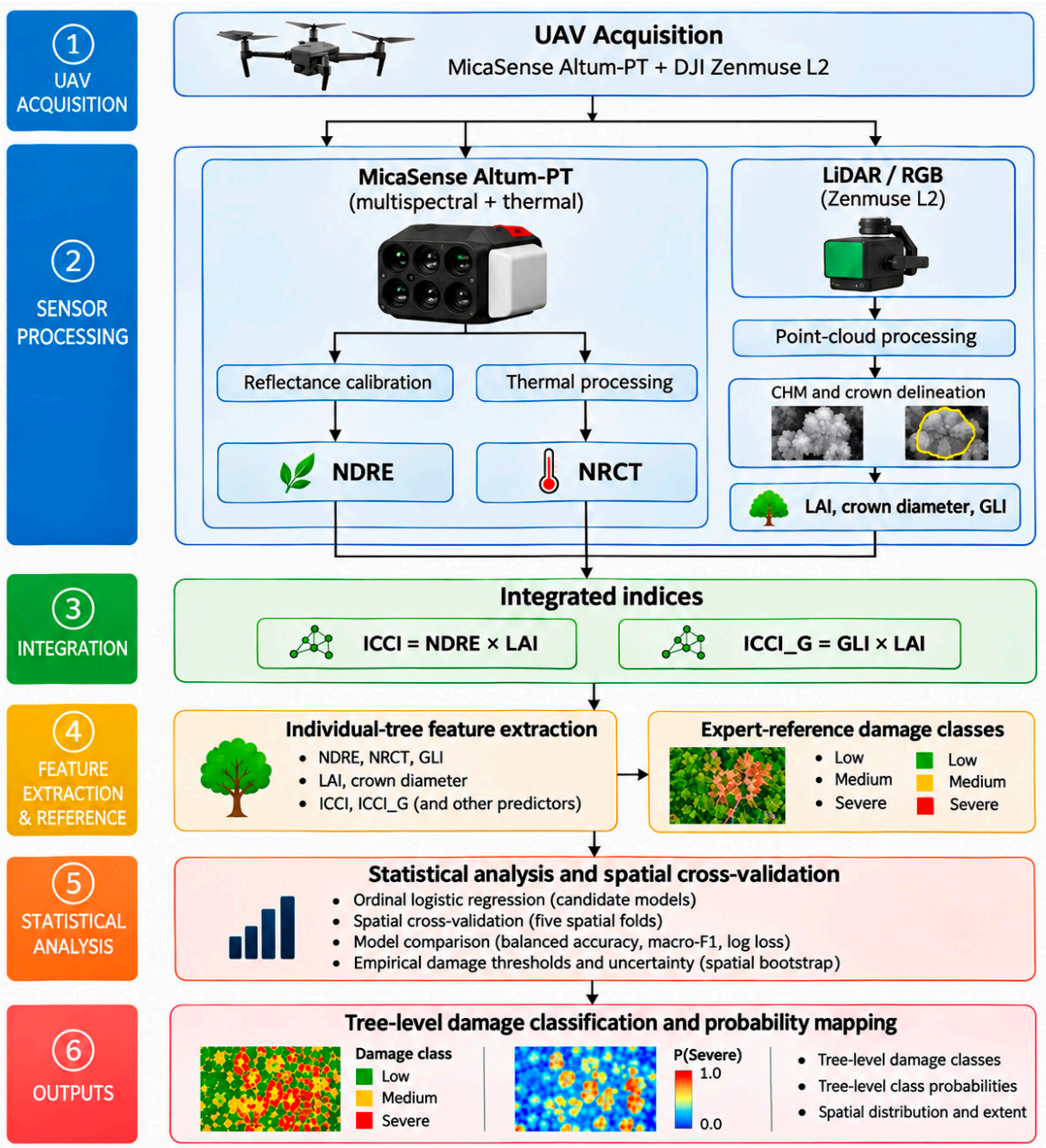

Figure 3. UAV-based workflow for quantitative post-hail damage assessment. Multispectral imagery provides spectral and thermal indicators of canopy condition, while ULS provides individual-tree and canopy structural attributes. The datasets are spatially integrated through an integrated canopy condition index (ICCI), and expert-defined damage zones provide reference observations for empirical calibration, ordinal damage classification, and spatial validation.

The resulting individual-tree inventory included tree location, height (HT), crown diameter (CD), crown area, diameter at breast height (DBH), and associated volume attributes. Crown diameter was calculated from segmented crown geometry, while tree height was determined from the maximum normalised canopy height associated with each crown. DBH was estimated from the adopted allometric relationship using tree height and crown dimensions [38].

3.3.2. Individual Tree Detection and Segmentation

Individual-tree delineation was undertaken from the LiDAR-derived canopy representation described in Section 3.3.1. Tree-top detections derived from variable-window

local maxima of the CHM were used as markers for crown segmentation. In addition, where stem-level point information was examined, a local density–times–intensity map (LDTIM) was used to enhance local stem signatures by combining point-density and relative-intensity information.

Point density was calculated from the number of returns occurring within each vertical raster column, while mean relative intensity was obtained from returns within the corresponding column. The normalised density and intensity layers were multiplied to form the LDTIM, which was smoothed using a 5×5 Gaussian kernel with $\sigma = 1$ (FWHM ≈ 2.36 pixels). LiDAR intensity was not radiometrically calibrated for range, incidence angle, or other acquisition effects and was therefore interpreted only in terms of relative local contrast rather than absolute reflectance.

Individual crowns were delineated using watershed segmentation [26,39]. The treetop detections were used as markers for segmentation of the CHM, after which crown boundaries were represented explicitly as polygons. Retaining the segmented crown geometry was important for subsequent analyses because the same crown polygons were used to extract spectral and structural observations and to quantify spatial overlap with the expert-assessed damage zones.

3.3.3. Structural Canopy Metrics

Return heights were normalised relative to the interpolated ground surface, with returns at least 1.0 m above ground classified as canopy. For each nominal 1 m grid cell, gap fraction (GF) was calculated as the proportion of recorded returns below this canopy-height threshold, equivalent to one minus the proportion classified as canopy:

$$\mathrm{GF} = 1 - \frac{N_{\mathrm{canopy}}}{N_{\mathrm{total}}},$$

where N_{canopy} is the number of returns classified as canopy and N_{total} is the total number of recorded returns within the cell. LAI was then estimated using

$$\mathrm{LAI} = -\frac{\cos\left(\overline{\theta}\right)}{0.5} \ln\left[\max\left(P_{\mathrm{gap}}, \epsilon\right)\right],$$

where $\overline{\theta}$ is the mean scan angle, 0.5 is the assumed projection/extinction coefficient, and ϵ is machine precision, used to avoid an undefined logarithm when the estimated gap fraction was zero. LAI was not capped at an upper limit. Crown-level LAI was calculated by averaging finite grid-cell values within the segmented crown polygons, with bilinear interpolation at the crown centroid used when no valid grid-cell centre fell within a polygon. This all-recorded-return formulation, designated E0 in the subsequent sensitivity analysis, was the original LiDAR-derived effective LAI estimator used to construct the structural predictors and integrated canopy condition indices.

This formulation treats all recorded returns as independent observations, including multiple returns originating from the same laser pulse. Consequently, the resulting LAI estimates are influenced by return multiplicity, foliage clumping, woody material, scan geometry, return density, and assumptions underlying the gap-fraction model [28,29]. Because foliage and woody components were not distinguished, and discrete-return penetration provides only an indirect measure of canopy interception, these estimates are more appropriately interpreted as LiDAR-derived effective structural interception or plant-area proxies rather than independently validated measurements of true foliage LAI.

CC was calculated as the proportion of recorded LiDAR returns classified as canopy within each analysis cell and was therefore the exact complement of the return-based GF, such that CC = 1 − GF. Accordingly, CC represents a canopy return fraction rather than an

independent estimate of projected canopy cover. Vertical point-density distributions and two-dimensional canopy profiles were additionally evaluated to characterise changes in crown form associated with needle and branch loss.

3.3.4. Multispectral Analysis

Multispectral observations were georeferenced to WGS84 and restricted to the spatial extent represented by the LiDAR data. The sensor dataset contained blue, green, red, red-edge, near-infrared (NIR), panchromatic, and long-wave infrared (LWIR) measurements, from which spatially continuous spectral and thermal point clouds were generated.

Vegetation indices were calculated to characterise complementary aspects of canopy vigour, chlorophyll content, moisture status, and vegetation stress. These included NDVI, NDRE, EVI, and GLI, together with additional vegetation and soil-adjusted indices evaluated during exploratory analysis. The principal spectral indices used in the integrated analysis were

$$\mathrm{NDRE} = \frac{\mathrm{NIR} - \mathrm{RedEdge}}{\mathrm{NIR} + \mathrm{RedEdge}}$$

and

$$\mathrm{GLI} = \frac{\mathrm{Green} - 0.5 * (\mathrm{Red} + \mathrm{Blue})}{\mathrm{Green} + 0.5 * (\mathrm{Red} + \mathrm{Blue})}$$

NDRE was selected as the principal red-edge indicator because of its sensitivity to canopy chlorophyll and reduced saturation relative to conventional NDVI under dense vegetation conditions. GLI was evaluated as a simpler alternative because it can potentially be derived without a dedicated red-edge band, providing a pathway towards lower-cost operational implementation using colourised LiDAR sensors.

Thermal observations were evaluated as complementary indicators of canopy stress and moisture condition using the normalised relative canopy temperature (NRCT),

$$\mathrm{NRCT} = \frac{T_c - T_{min}}{T_{max} - T_{min}},$$

where T_c is the canopy temperature and T_{min} and T_{max} the minimum and maximum valid temperatures, respectively, used to define the normalisation range. Thermal values were converted from the Agisoft Metashape digital number (DN) export using $T = \frac{DN}{100} - 273.15$, with no additional radiometric or emissivity correction applied.

The thermal observations were spatially aggregated to a 1 m raster, and NRCT was calculated using the absolute minimum and maximum temperatures across all valid raster cells within the plantation/LiDAR analysis extent; these extrema were −8.4 °C and 52.5 °C, respectively, and were applied consistently across the study area. No canopy or vegetation mask was applied when determining the normalisation bounds. Consequently, the extrema could include observations of canopy, exposed soil, understory, and other surfaces. NRCT values were subsequently extracted and averaged within the segmented individual-tree crown polygons, again with no additional radiometric or emissivity correction applied. NRCT therefore represents relative thermal variation within the surveyed area rather than an independently calibrated measure of canopy physiological temperature.

Multispectral measurements were aggregated to a 1 m spatial grid to provide a common analysis framework with the LiDAR-derived canopy variables. This resolution represented a compromise between retaining within-crown spectral information and reducing differences in spatial support among the sensor products. Given that individual crowns were typically approximately 3 m in diameter, a sensitivity analysis was also undertaken using finer-resolution NDRE measurements and crown-interior extraction to assess the potential influence of spatial aggregation and mixed canopy–background pixels on crown-level spectral estimates.

3.3.5. Spatial Integration and Integrated Canopy Condition Index

LiDAR and multispectral observations were transformed to the same coordinate reference system and mapped to a common 1 m analysis grid, ensuring that structural and spectral measurements represented coincident ground locations. Spatial extents and overlap were recorded as quality-control measures prior to integration.

To integrate physiological condition and canopy structure within a single metric, the integrated canopy condition index (ICCI) was defined as ICCI = LAI × NDRE. A second formulation, $ICCI_G$, was defined by replacing NDRE with GLI. The multiplicative formulation was adopted as a simple empirical composite in which the index decreases when either LiDAR-derived foliage density or spectral canopy condition declines. It was intended to provide an interpretable means of combining complementary structural and spectral observations rather than to represent a mechanistic model of canopy physiology.

The utility of the composite indices was evaluated against their constituent measurements and the expert-reference damage classes. The green leaf index (GLI) was calculated as $(2G - R - B)/(2G + R + B)$, where R, G, and B represent the red, green, and blue image channels, respectively. The composite indices ICCI (NDRE × LAI) and $ICCI_G$ (GLI × LAI) were evaluated as standalone predictors rather than as statistical interaction terms within the regression models.

ICCI combines the spectral response of the remaining needle foliage with LiDAR-derived effective canopy foliage density, allowing physiological and structural consequences of hail damage to be represented within a common index. Because GLI can be derived from RGB information associated with the LiDAR acquisition, $ICCI_G$ was evaluated as a simpler operational alternative where red-edge observations are unavailable or registration of independently acquired multispectral and LiDAR datasets is undesirable. The RGB information used to derive GLI in this study was obtained from the Zenmuse L2 colour information rather than the separately acquired Zenmuse P1 imagery.

3.3.6. Expert Reference Damage Assignment

Expert assessment of hail-induced crown damage was undertaken by a single trained forestry assessor with approximately 20 years of experience in plantation forest assessment, including five years of experience in forest-health aerial assessment involving damage agents. The assessor manually delineated polygons representing low, medium, and severe crown-damage zones using a geographic information system (GIS) and the DJI P1 RGB orthomosaic acquired for this study on 8 November 2024. The polygons were prepared specifically for the research, without automated classification or segmentation.

Damage severity was assessed primarily from visible needle loss and crown thinning attributable to the hailstorm, using a defoliation-based approach consistent with established forest-health assessment practices outlined in [40,41]. The underlying severity-rating scheme distinguished trace (>0–10% defoliation), low (>10–20%), medium (>20–40%), high (>40–50%), and severe (>50%). However, no areas were assigned to the trace or high categories. Consequently, the expert-reference classification comprised three observed severity classes: low, medium, and severe.

Classification was based on the assessor's professional judgement of the predominant extent and severity of hail-induced crown damage within each mapped area. Damage extent referred to the spatial area affected by hail, whereas severity described the estimated proportion of foliage lost due to the storm. The assessment focused on these characteristics rather than estimating the incidence of damage among individual trees.

Individual-tree characteristics other than defoliation, such as branch loss, stem injury, or physiological status were not used as they were not visible. Where damage severity varied spatially, the predominant severity was assigned. In areas where the boundary

between adjacent severity classes was ambiguous, the assessor adopted a conservative approach, assigning the lower damage-severity rating rather than the higher rating. There were no zones of trace or high levels of damage. Small, isolated patches and individual trees exhibiting different levels of damage were not separately delineated.

The assessor used the DJI P1 RGB orthomosaic but did not use MicaSense Altum-PT multispectral imagery, vegetation-index maps, thermal imagery, LiDAR-derived structural measurements, or the subsequently developed model predictions when delineating the reference polygons. The reference assessment was therefore conducted independently of the spectral, thermal, and structural predictor variables evaluated in the classification models. The polygons were not subsequently modified in response to the remote-sensing analyses or model predictions.

The assessor also undertook field inspections on 29 October and 6 December 2024, including vehicle-based inspections along compartment breaks and approximately ten walking inspections into the plantation. Areas representing all damage-severity classes were visited, and supporting information included georeferenced photographs, GPS observations, and operational inspection reports. These observations provided ground-based corroboration of the expert-delineated damage zones but were not used to modify polygon boundaries or assigned severity classes. The inspections did not constitute a systematic, independently labelled sample of individual trees covering all severity classes.

The resulting polygons represented areas of predominant expert-assessed crown-damage severity, delineated using conventional manual digital sketch-mapping techniques. Although each polygon was assigned a single predominant severity class, individual trees within a polygon could exhibit different levels of damage. The polygons were therefore treated as an expert-reference spatial classification against which remotely sensed observations could be evaluated, rather than as independently verified individual-tree damage assessments or error-free biological ground truth.

For each segmented crown, the proportional area intersecting each of the three expert-defined damage zones was calculated, and the zone with the greatest proportional overlap was designated the dominant reference class. A crown was assigned to that class only when at least 70% of its area occurred within the dominant zone and it was not classified as a boundary crown. Boundary crowns were defined as crowns intersecting more than one expert zone or crowns whose apex-to-boundary distance was no greater than the larger of the crown radius and 1.0 m. These combined reference-quality criteria left 151,139 assigned crowns. A further 117 crowns lacking complete predictor data were removed, producing the final analytical population of 151,022 trees.

For each accepted tree crown, distributional statistics were extracted from the NDRE, GLI, NRCT, LAI, ICCI, and $ICCI_G$ rasters, including the mean, median, standard deviation, quartiles, 90th percentile, minimum, and maximum. These crown-level observations were combined with LiDAR-derived structural attributes to form the dataset used for statistical calibration and damage modelling. Consequently, subsequent classification-performance statistics quantify agreement with polygon-derived expert-reference classes rather than independently validated accuracy in detecting the actual severity of hail damage to individual trees.

3.3.7. Statistical Analysis and Empirical Damage Classification

Differences among expert-assigned low, medium, and severe damage classes were first examined independently for each candidate structural, spectral, thermal, and integrated canopy metric. Descriptive statistics were calculated for each class, and differences were evaluated using Kruskal–Wallis tests, ANOVA diagnostics, pairwise comparisons, and corre-

sponding effect-size measures. These analyses were used to characterise the discriminatory capacity of individual metrics across independently assessed levels of hail damage.

Empirical classification thresholds were then derived for individual damage metrics at two ordinal transitions: low versus medium-or-severe and low-or-medium versus severe. Thresholds were selected by maximising Youden's J statistic [42],

$$J = \text{Sensitivity} + \text{Specificity} - 1$$

Full-data and cross-validated thresholds were calculated, and threshold stability was assessed using spatial-cluster bootstrap resampling (hereafter, spatial bootstrap). Because neighbouring trees were spatially correlated, predictive performance was evaluated using spatially separated rather than randomly allocated validation folds, thereby reducing the risk of optimistic performance estimates arising from spatial dependence [43–45].

Ten irregular spatial clusters were defined by k-means clustering of the projected tree coordinates (X, Y). Rather than resampling individual trees independently, these clusters were sampled with replacement, with all eligible trees belonging to each selected cluster retained together. For each of 2000 bootstrap iterations, empirical thresholds were re-estimated independently for each candidate metric at both ordinal transitions. The resulting bootstrap distributions were used to derive 95% bootstrap intervals and assess threshold stability.

Damage severity was subsequently modelled as an ordered categorical response using proportional-odds ordinal logistic regression. Candidate models incorporating different combinations of spectral, thermal, structural, and integrated canopy condition predictors were evaluated (Table 1).

Table 1. Candidate proportional-odds ordinal logistic regression models evaluated for classification of expert-assigned low, medium, and severe hail damage using spectral, thermal, structural and integrated canopy condition predictors. For multivariable models, listed variables entered the regression as separate predictors.

Model	Predictors
M1	NDRE
M2	LAI
M3	CD
M4	ICCI
M5	$ICCI_G$
M6	LAI + CD
M7	NDRE + LAI + CD
M8	ICCI + CD
M9	$ICCI_G$ + CD
M10	NDRE + LAI + CD + $ICCI_G$
M11	NRCT
M12	NDRE + NRCT
M13	GLI
M14	GLI + LAI
M15	NDRE + LAI

To assess whether incorporating LiDAR-derived LAI provided additional discriminatory information beyond spectral indices alone, three supplementary candidate models were evaluated: GLI alone (M13), GLI and LAI as separate predictors (M14), and NDRE and LAI as separate predictors (M15). These models were compared with the existing NDRE-only (M1), ICCI (M4) and $ICCI_G$ (M5) formulations. All models were evaluated using the same 151,022 reference trees and identical five-fold spatial cross-validation assignments. Predictor centring and scaling parameters were estimated within each training fold

and applied to the corresponding held-out observations. Paired differences in balanced accuracy, macro-F1 and log loss were calculated using pooled out-of-fold predictions, with uncertainty estimated from 2000 paired bootstrap resamples of ten spatial clusters. These bootstrap intervals characterise spatial variability in differences between fixed out-of-fold predictions; models were not refitted during resampling.

The proportional-odds assumption was assessed using the conventional Brant test of parallel slopes [46]. For the three ordered damage classes, separate binary logistic regressions were fitted at the two cumulative boundaries (low vs. medium-or-severe and low-or-medium vs. severe), using the complete 151,022-tree reference dataset. NDRE and NRCT were standardised using the full-data means and standard deviations. The test evaluates whether the predictor coefficients are equal across the two cumulative boundaries, using a Wald statistic incorporating the model-implied covariance between the nested binary responses. Global and predictor-specific tests were performed with two and one degrees of freedom, respectively. Because the conventional Brant test treats individual trees as statistically independent, its significance was interpreted cautiously in the presence of spatial autocorrelation. The practical consequences of departures from proportional odds were therefore examined separately through spatially cross-validated comparisons with partial proportional-odds and multinomial logistic regression models.

To assess the practical consequences of departures from proportional odds in M12 (NDRE + NRCT), additional model specifications were evaluated. Two partial proportional-odds (PPO) specifications were evaluated: one allowing the NRCT coefficient to vary between ordinal boundaries while retaining a common NDRE coefficient (PPO-1), and another allowing both coefficients to vary (PPO-2). A multinomial logistic regression model, which does not impose proportional odds or explicitly model the ordering of damage classes, was also fitted. The PPO implementations incorporated constraints to ensure valid, non-crossing cumulative probabilities over the specified predictor domain. These alternative specifications were used to assess whether relaxing proportional odds materially affected model interpretation, classification performance or predicted probabilities.

The integrated indices ICCI and $ICCI_G$ were calculated as described in Section 3.3.5, while NRCT was derived from canopy temperature as described in Section 3.3.4. For multivariable models, the variables listed in Table 1 entered the regression as separate predictors rather than being combined arithmetically. Predictor dependence was evaluated using Pearson and Spearman correlations, variance inflation factors, and condition diagnostics.

Within each spatial cross-validation training fold, continuous predictors were standardised using the corresponding training-fold mean and standard deviation. For tree i, the linear predictor for each candidate model was defined as

$$\eta_i = \sum_{k=1} \beta_k \, z_{k_i},$$

where z_{k_i} is the standardised value of predictor for tree i, β_k is its estimated regression coefficient, and p is the number of predictors included in the candidate model. Regression coefficients and ordinal cut-points were estimated jointly by maximum likelihood.

For the three ordered damage classes, the proportional-odds model was expressed as

$$\log\left[\frac{P(Yi \leq j)}{P(Yi > j)}\right] = \alpha_j + \eta_i, \quad j = 1, 2$$

where α_j represents the class-specific cut-point. The same regression coefficients were applied at both ordinal boundaries under the proportional-odds assumption.

With damage severity ordered as low < medium < severe, the proportional-odds model was parameterised using cumulative logits, where each ordinal boundary contrasted the probability of being at or below that boundary with the probability of exceeding it.

Because predictors were standardised, each regression coefficient represents the change in cumulative log-odds associated with a one-standard-deviation increase in the corresponding predictor, conditional on the other predictors in the model. A positive coefficient therefore indicates increased odds of being at or below the specified severity boundary and is associated with lower damage severity, whereas a negative coefficient indicates a shift towards higher damage severity. Under the proportional-odds assumption, the same coefficient applies at both ordinal boundaries. Where this assumption is violated, the common coefficients should be interpreted as constrained summary associations rather than evidence of identical predictor effects across severity boundaries.

Model performance was principally assessed using five-fold spatial cross-validation to reduce the optimistic bias that would arise from randomly partitioning neighbouring trees between training and validation datasets. Trees were partitioned according to their projected spatial coordinates using k-means clustering, producing five geographically structured folds; each fold was withheld in turn while the model was fitted to the remaining four folds. Predictor means and standard deviations were calculated exclusively from the training data in each fold and subsequently applied to the held-out observations. Candidate models were compared using balanced accuracy, macro-F1 score, overall accuracy, and stability among spatial validation folds.

The use of 10 bootstrap clusters and five cross-validation folds was pre-specified rather than determined through formal optimisation; the different numbers of partitions provided multiple spatial resampling units for bootstrap uncertainty estimation while maintaining sufficiently large spatially partitioned training and validation samples for model assessment.

Following model selection, the selected model was refitted to the complete reference dataset and used to generate probabilities of low, medium, and severe damage for each tree, from which the final predicted class was assigned.

Spatial uncertainty in the fitted ordinal coefficients was additionally evaluated using the same 10-cluster bootstrap framework described above. For each bootstrap replicate, predictors were re-standardised using the resampled observations before refitting the proportional-odds model. The resulting coefficient distributions were used to derive spatial-bootstrap uncertainty intervals.

3.3.8. Model Validation and Sensitivity Analysis

Predictive performance was evaluated using confusion matrices, overall and balanced accuracy, macro-precision, macro-recall, macro-F1 score, class-specific sensitivity and specificity, predictive values, and one-versus-rest receiver operating characteristic (ROC) curves with associated area-under-the-curve (AUC) statistics. Spatially blocked cross-validation provided the principal assessment of predictive generalisation.

Uncertainty in model calibration and empirical thresholds was evaluated using the spatial bootstrap described in Section 3.3.7. Conventional stratified bootstrap estimates were retained as secondary diagnostics.

Sensitivity analyses examined the influence of the minimum crown overlap required for expert assignment, exclusion buffers around expert-zone boundaries, perturbations of empirical thresholds, alternative crown-level raster summary statistics, and alternative formulations of ICCI and $ICCI_G$.

The sensitivity of crown-level NDRE measurements to spatial aggregation and crown-boundary effects was evaluated by comparing the original 1 m grid with a separately aggregated 0.25 m grid derived from the original calibrated Metashape spectral point export. Point-level NDRE values were averaged within nominal 0.25 m cells before extraction using the original segmented crown polygons. The finer grid was not produced by interpolation or up-sampling of the existing 1 m raster. Because the Metashape export did not retain

source-image identifiers or the number of overlapping photographs contributing to each reconstructed point, the observations cannot be considered statistically independent ground-resolution measurements. Accordingly, 0.25 m refers to the aggregation-cell size rather than the verified native or effective spatial resolution of the imagery. Additional analyses applied inward polygon buffers of 0.25 m and 0.50 m to exclude observations near crown boundaries. The original expert-reference assignments were retained, and comparisons were restricted to 145,823 trees with valid measurements under all four configurations. Crown-level NDRE distributions, agreement statistics, empirical thresholds, severe-damage AUC, and spatially cross-validated classification performance were compared using the original five spatial folds. Uncertainty in paired performance differences was estimated using 2000 bootstrap resamples of ten spatial clusters.

As an additional sensitivity analysis, spatial validation was repeated using five contiguous bands defined by quantiles of projected X coordinate. This directional partitioning provided stronger separation along the principal west–east damage gradient than the primary k-means spatial cross-validation. Each band was withheld in turn, with predictor standardisation derived exclusively from the remaining training observations. The analysis was restricted to the NDRE-only model (M1) and selected NDRE–NRCT model (M12). Two additional edge-to-edge tests were performed by training on the western 80% of the plantation and predicting the eastern 20%, and vice versa. Because damage classes were strongly spatially segregated, class representation within each directional holdout was examined explicitly and performance measures from class-incomplete holdouts were interpreted as sensitivity diagnostics rather than directly comparable estimates of three-class predictive accuracy. Predictor-range overlap between training and validation observations was also quantified to distinguish spatial transfer from extrapolation beyond the observed predictor ranges.

The sensitivity of LiDAR-derived LAI to return-selection and gap-fraction assumptions was evaluated using three estimators: the all-return formulation (E0), an otherwise equivalent formulation restricted to first returns (E1), and a recorded-pulse GF formulation (E2). For E2, individual recorded pulses were reconstructed using GPS-time groupings within the original LAS files, and each complete pulse was treated as a single observation. Pulse groups lacking a unique first return or the declared number of returns were excluded. Because the available data did not identify emitted pulses that produced no recorded return, E2 was interpreted as a recorded-pulse proxy rather than a fully emission-normalised LAI estimator. Only 31.06% of reconstructed recorded-pulse groups satisfied the conservative completeness criteria.

The proportional-odds sensitivity analysis compared the original M12 model with PPO-1, PPO-2, and multinomial logistic regression using the same 151,022-tree reference population and original five spatial cross-validation folds. Predictor standardisation was performed independently within each training fold, and withheld observations were assigned to the class with the highest predicted probability. Model performance was assessed using balanced accuracy, macro-F1, class-specific recall, multiclass log loss, and the multiclass Brier score. Probability calibration was additionally examined using class-specific reliability diagrams and expected calibration error. Paired differences between models were evaluated using 2000 bootstrap resamples of ten spatial clusters applied to the fixed out-of-fold predictions. The resulting confidence intervals quantify spatial variability conditional on the fitted models and do not incorporate uncertainty associated with model refitting.

The partial proportional-odds models incorporated constraints to ensure coherent cumulative probabilities across the predictor domain. These constraints were transformed using the training-fold means and standard deviations to maintain consistency with predictor standardisation. The primary implementation imposed constraints over the conservative

two-dimensional domain $[0,1]^2$, encompassing the observed NDRE values (0.056–0.784) and the normalised NRCT range. Because NDRE is theoretically bounded by $[-1, 1]$, an additional sensitivity analysis examined the effect of restricting the constraints to the observed predictor support, thereby allowing greater flexibility in the boundary-specific coefficients.

Additional analyses examined the original all-return gap fraction and the quantity previously designated canopy cover. Code inspection established that the latter represented the proportion of recorded returns classified as canopy, making it the exact complement of gap fraction rather than an independent measure of geometric canopy cover. Accordingly, only gap fraction was retained for comparative modelling. Sensitivity to canopy-height thresholds of 0.5, 1.0, and 1.5 m was also examined.

Estimator distributions, crown-level agreement, damage-class separation, and classification performance were compared using the original expert-reference assignments and spatial cross-validation framework. Classification comparisons used common-tree populations, the original five spatial folds, and training-fold-only predictor standardisation. Paired uncertainty was evaluated using 2000 bootstrap resamples of ten spatial clusters applied to fixed out-of-fold predictions. These intervals describe spatial variability conditional on the fitted models and do not incorporate model-refitting uncertainty.

The robustness of NRCT was assessed by comparing the original survey-wide absolute temperature extrema with alternative canopy-restricted and percentile-based normalisation bounds. Non-clipped alternatives were evaluated to determine whether affine rescaling affected crown-level thermal discrimination or refitted ordinal classification predictions. The influence of non-canopy observations was examined separately by repeating crown-level NRCT extraction using LiDAR-derived structural masks with LAI thresholds of >0 and >0.5, while retaining the original globally normalised thermal raster and segmented crown polygons. The baseline crown-level NRCT extraction was first reproduced exactly for all 151,022 reference trees. Classification comparisons were then conducted on a common population of 148,730 trees with valid measurements under all principal configurations, using the original five spatial cross-validation folds, training-fold-only standardisation, and ordinal logistic regression models M1 (NDRE) and M12 (NDRE + NRCT). Paired performance differences were evaluated using 2000 bootstrap resamples of ten spatial clusters.

4. Results

4.1. Expert-Derived Damage Classes and Tree Assignment

Expert interpretation delineated three spatially coherent levels of hail damage across the study plantation (Figure 4). The mapped areas of the low, medium, and severe damage zones were approximately 38.6 ha (20.9%), 58.7 ha (31.9%), and 87.0 ha (47.2%), respectively. The damage pattern exhibited a pronounced spatial gradient, with low-damage forest in the west transitioning through an intermediate zone to an extensive severely damaged area in the east. The irregular boundaries between zones indicate that the transition in damage severity was spatially heterogeneous rather than occurring as a simple linear front. All reported spatial results are presented in local east–north–up (ENU) coordinates, with horizontal positions expressed as local easting and northing in metres.

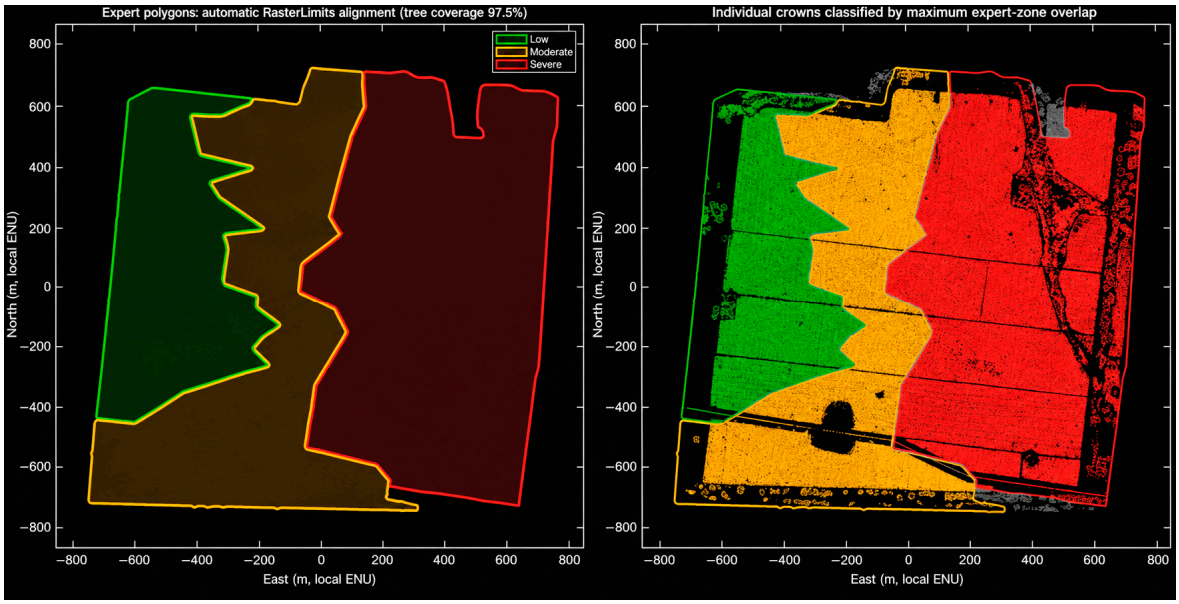

Figure 4. Expert-reference classification of hail damage across the study site. (**left**) Expert-defined zones representing low, medium, and severe hail damage and (**right**) corresponding assignment of segmented individual-tree crowns to the expert-reference classes based on spatial overlap. Coordinates are expressed in a local coordinate system.

Following spatial intersection of the expert-reference polygons with the remotely sensed tree data and application of the crown-assignment and exclusion criteria described in Section 3.3.6, 151,022 trees remained for analysis. Of these, 29,228 (19.4%) were assigned to the low-damage class, 49,140 (32.5%) to medium damage and 72,654 (48.1%) to severe damage. The resulting tree-level reference dataset therefore retained only crowns that could be assigned unambiguously to an expert-defined damage class and for which complete predictor data were available. These assignments should be interpreted as tree-level representations of the spatial expert crown-damage classification rather than independent field measurements of damage to each tree.

4.2. Spectral and Structural Responses to Damage Severity

Clear differences among expert-identified damage classes were observed in several remotely sensed canopy metrics (Figure 5). The strongest separation occurred for NDRE, which decreased progressively with increasing damage severity. Mean NDRE declined from 0.651 in the low-damage class to 0.593 for medium damage and 0.398 for severe damage. Corresponding median values were 0.661, 0.610, and 0.394. The severe class was therefore strongly displaced towards lower NDRE values, although some overlap remained between adjacent classes.

The integrated canopy indices showed clear monotonic responses to damage. Mean ICCI decreased from 2.876 for low-damage trees to 2.279 for medium damage and 1.380 for severe damage. The RGB-derived $ICCI_G$ exhibited an even stronger relative reduction, with class means of 0.697, 0.491, and 0.131 for low, medium, and severe damage, respectively. Median $ICCI_G$ values similarly declined from 0.684 to 0.494 and 0.098. Thus, both integrated indices reproduced the expected decline in canopy condition with increasing expert-assessed hail damage. Overall, separation among expert damage classes was greatest for NDRE and the integrated indices, whereas LAI and crown diameter showed substantially greater inter-class overlap.

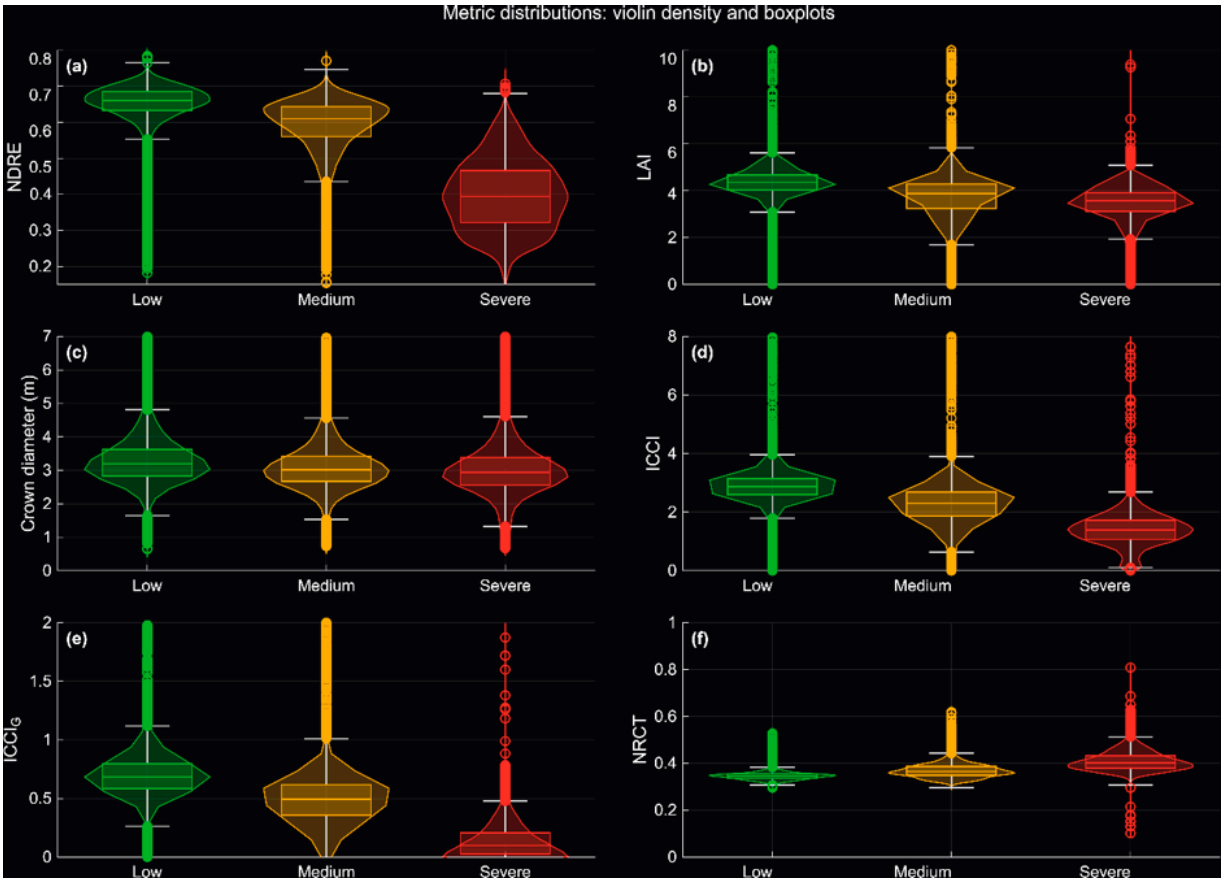

Figure 5. Distributions of canopy metrics against expert-assigned hail-damage classes. Violin plots and superimposed boxplots showing the distributions of (**a**) normalised difference red-edge index (NDRE), (**b**) LiDAR-derived effective leaf area index (LAI), (**c**) crown diameter (CD), (**d**) integrated canopy condition index (ICCI), (**e**) green-based integrated canopy condition index (ICCI$_G$), and (**f**) normalised relative canopy temperature (NRCT) for individual trees assigned to low (green), medium (amber), and severe (red) damage classes. Violin widths represent the estimated probability density, while the superimposed boxplots indicate the median and interquartile range. Individual markers represent observations beyond the boxplot whiskers. Axis ranges are restricted where appropriate to facilitate visualisation of the principal distributions; extreme observations remain included in all statistical analyses.

4.3. Empirical Discrimination and Damage Thresholds

ROC analysis was used to derive empirical thresholds for the two ordinal transitions: low versus medium-or-severe damage, and low-or-medium versus severe damage. NDRE provided strong discrimination for both transitions. The optimum thresholds were 0.606 for separating low damage from the combined medium/severe classes and 0.534 for separating severe damage from the combined low/medium classes. The corresponding AUC values were 0.899 and 0.954, respectively. For the severe-damage transition, sensitivity and specificity were 0.904 and 0.877, giving a Youden index of 0.781.

The structural variables were considerably weaker discriminators. LAI produced AUC values of 0.804 and 0.714 for the two transitions, with thresholds of approximately 4.00 in both cases. Crown diameter produced AUC values of only 0.602 and 0.569, confirming the substantial overlap evident in its class distributions.

ICCI thresholds were 2.342 for the low versus medium-or-severe transition and 1.984 for the low-or-medium versus severe transition. The corresponding AUC values were 0.887 and 0.897. The equivalent ICCI$_G$ thresholds were approximately 0.492 and 0.334. Among the evaluated metrics, NDRE produced the highest AUC for the severe-damage transition.

Spatial bootstrap analysis indicated that the empirical thresholds were generally reproducible under spatial resampling (Figure 6).

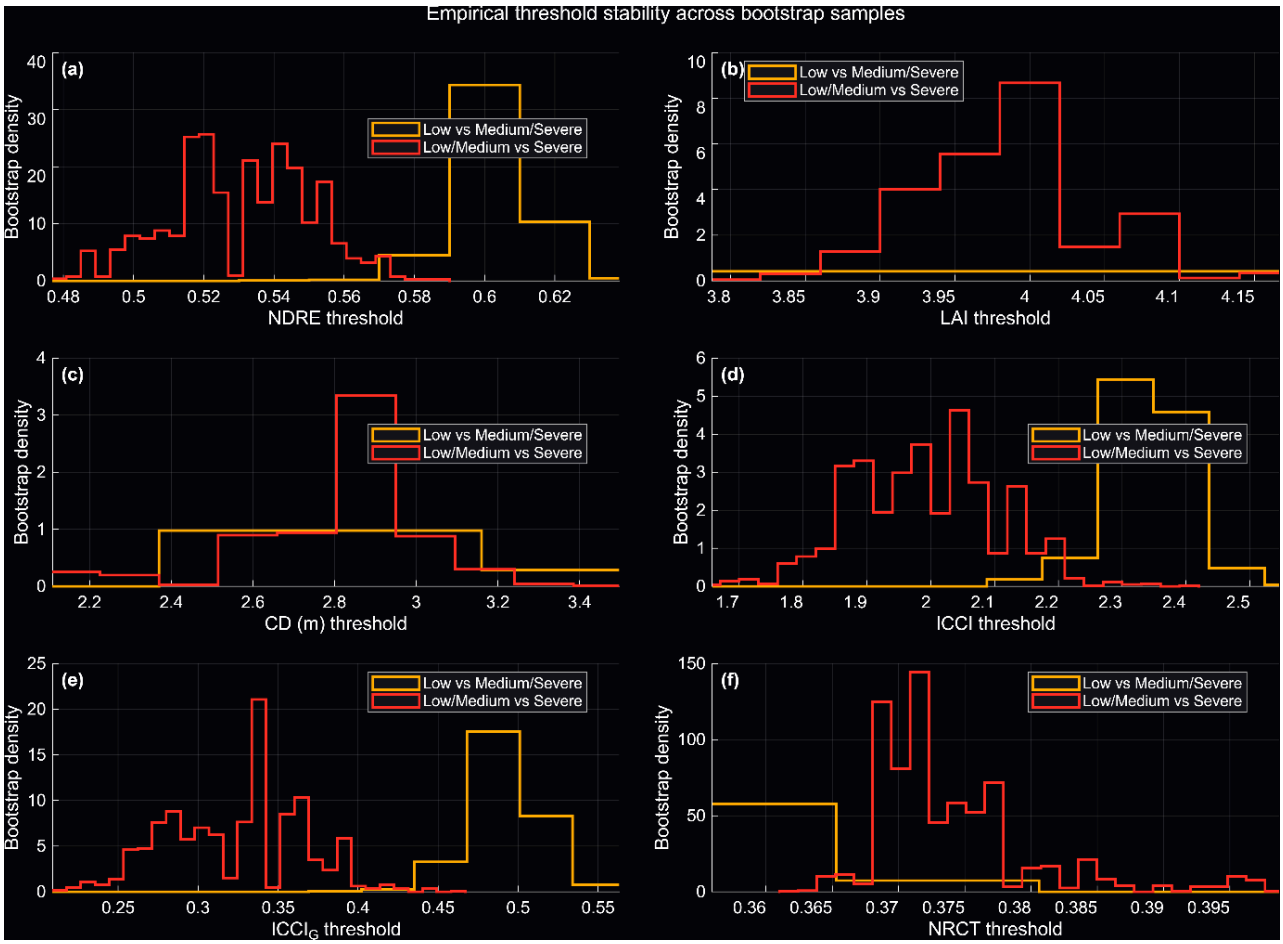

Figure 6. Spatial bootstrap stability of empirical hail-damage thresholds. Bootstrap distributions of empirically derived thresholds separating low damage from medium-or-severe damage and low-or-medium damage from severe damage for (**a**) NDRE, (**b**) leaf area index (LAI), (**c**) crown diameter, (**d**) integrated canopy condition index (ICCI), (**e**) the RGB-based integrated canopy condition index (ICCIG), and (**f**) normalised relative canopy temperature (NRCT). Threshold uncertainty was evaluated using 2000 spatial cluster (block) bootstrap iterations. Narrower distributions indicate greater stability of the corresponding class boundary.

Across 2000 bootstrap iterations, mean NDRE thresholds were 0.601 (95% bootstrap interval: 0.573–0.625) for low versus medium-or-severe damage and 0.530 (0.489–0.568) for low-or-medium versus severe damage; corresponding mean ICCI thresholds were 2.344 (2.185–2.454) and 1.990 (1.777–2.201), while $ICCI_G$ thresholds averaged 0.490 (0.451–0.538) and 0.323 (0.253–0.395), respectively. The comparatively concentrated bootstrap distributions indicate that the empirical class boundaries for NDRE and the integrated indices were reasonably stable to changes in the spatial composition of the calibration sample, although these intervals represent within-site spatial uncertainty rather than evidence of threshold transferability to independent plantations.

4.4. Performance of the Integrated Canopy Indices

Initial comparisons of the integrated canopy indices were undertaken using proportional-odds ordinal logistic regression models fitted to the complete expert-reference dataset. These analyses produced apparent (re-substitution) balanced accuracies of 0.650 for ICCI and

0.691 for $ICCI_G$, compared with 0.707 for NDRE alone (Table 2). Although $ICCI_G$ exhibited higher apparent classification performance than ICCI, these values were calculated from predictions for the same observations used for model fitting and should not be interpreted as spatially independent estimates of predictive performance.

Table 2. Apparent (re-substitution) classification performance of univariate proportional-odds ordinal logistic regression models fitted to the complete expert-reference dataset. Predictors comprise mean crown NDRE, mean LAI, crown diameter, ICCI and ICCI_G. Performance metrics were calculated using predictions for the same trees employed in model fitting and therefore represent apparent agreement rather than spatially independent predictive performance. Weighted κ accounts for the ordered nature of the three damage classes, while mean absolute class error represents the average ordinal distance between observed and predicted classes. Spatially cross-validated model comparisons are reported separately and provide the principal basis for evaluating predictive performance.

Metric	Mean NDRE	Mean LAI	Crown Diameter (CD)	Mean ICCI	Mean $ICCI_G$
Accuracy	0.755	0.542	0.476	0.702	0.742
Balanced accuracy	0.707	0.423	0.332	0.650	0.691
Weighted κ	0.766	0.343	−0.001	0.707	0.750
Mean absolute class error	0.252	0.529	0.717	0.309	0.268
Log loss	0.563	0.982	1.029	0.754	0.636

Under five-fold spatial cross-validation, balanced accuracy decreased to 0.588 for ICCI and 0.644 for $ICCI_G$, compared with 0.651 for NDRE alone. Thus, $ICCI_G$ retained an advantage over ICCI under spatial validation but did not outperform the standalone NDRE predictor. Additional comparisons demonstrated that GLI alone achieved a balanced accuracy of 0.636, only 0.008 lower than $ICCI_G$. This difference was uncertain under paired spatial-cluster bootstrap resampling (95% interval: −0.022 to 0.037). Incorporating LAI as an independent predictor alongside GLI produced negligible improvement (balanced accuracy 0.636), as did adding LAI to NDRE (0.650 compared with 0.651 for NDRE alone). These results indicate that the apparent superiority of $ICCI_G$ over ICCI does not establish a general advantage from integrating LiDAR-derived LAI with spectral measurements.

The full-data results in Table 2 are retained as descriptive measures of apparent model agreement with the expert-reference classifications. Spatially cross-validated results provide the principal basis for comparing predictive performance among candidate models.

4.5. Sensitivity of Empirical Classification Thresholds

Classification performance varied systematically when the empirically derived thresholds were perturbed (Figure 7). For NDRE, the nominal thresholds of 0.606 and 0.534 produced an overall classification accuracy of 0.702. Increasing both thresholds by 5% raised accuracy to 0.736, the maximum observed within the tested perturbation range, after which performance declined. Negative perturbations progressively reduced accuracy. Accuracy at +10% remained slightly above the nominal value, whereas larger positive departures reduced performance below the nominal level.

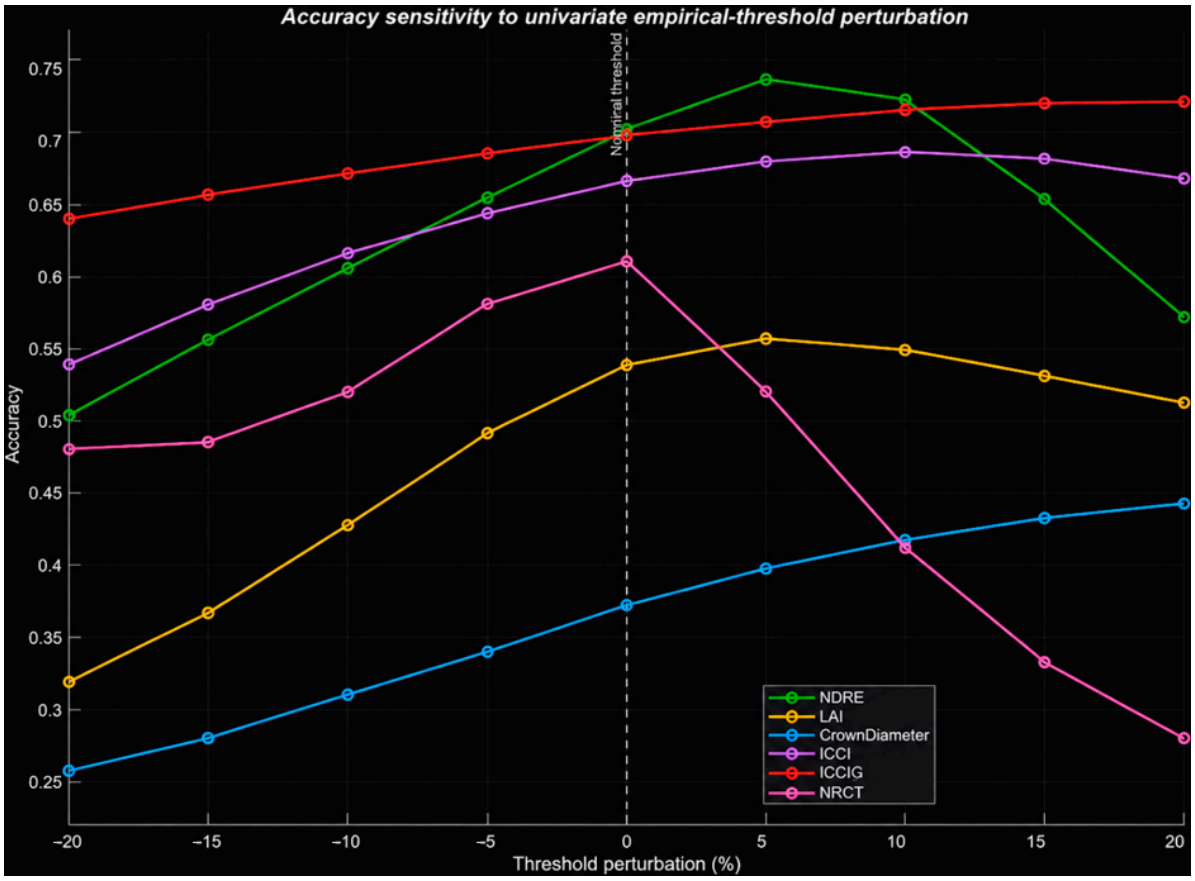

Figure 7. Sensitivity of empirical hail-damage classification accuracy to perturbations of the nominal thresholds for NDRE, LAI, crown diameter, ICCI, ICCI_G, and NRCT. Both thresholds defining the three damage classes were varied simultaneously from −20% to +20% of their nominal values. The dashed vertical line indicates the original empirically derived thresholds (0% perturbation). Accuracy was calculated by applying deterministic threshold-based classification rules to the analysed tree population and should not be interpreted as spatially cross-validated predictive performance. The nominal thresholds were derived from separate binary ROC analyses and therefore do not necessarily maximise overall three-class classification accuracy.

The nominal-threshold accuracy of 0.702 should not be conflated with the apparent accuracy of 0.755 reported for the NDRE ordinal model in Table 2. The former was obtained by applying the two empirically derived NDRE thresholds as a deterministic classification rule, whereas the latter was obtained from re-substitution predictions of a continuous-predictor proportional-odds ordinal model fitted to the complete eligible reference dataset. Neither value represents spatially independent validation performance.

At the nominal NDRE thresholds, 54,664 trees were classified as having low damage, 23,685 as medium damage, and 77,793 as severe damage; these classes represented approximately 35.0%, 15.2%, and 49.8% of the classified trees, respectively. Severe-damage recall at the nominal thresholds was 0.904.

The other candidate metrics exhibited distinct sensitivity profiles. ICCI showed a comparatively broad response, with accuracy increasing gradually from approximately 0.67 at the nominal thresholds to a maximum of about 0.69 at a +10% perturbation. $ICCI_G$ was comparatively insensitive to positive threshold perturbation, increasing from approximately 0.70 at the nominal thresholds to about 0.73 at the upper end of the tested range. In contrast, NRCT showed a more pronounced peak near its nominal thresholds, with accuracy declining substantially under larger positive perturbations. These results reinforce that the

empirical thresholds should be interpreted as statistical decision points rather than fixed biological boundaries.

Sensitivity analysis of the expert-reference assignment showed that classification accuracy was effectively invariant to the minimum crown–polygon overlap criterion across the tested range of 50–90% but increased modestly as trees close to expert-zone boundaries were excluded. Accuracy increased from approximately 0.755 with no additional boundary buffer to 0.769 with a 10-m buffer (Figure 8), while the nominal reference-assignment procedure produced an accuracy of approximately 0.757. These results indicate that classification performance was robust to the crown-overlap criterion and only modestly sensitive to boundary exclusion, with the progressive increase in accuracy consistent with greater reference uncertainty among trees located near expert-defined class boundaries.

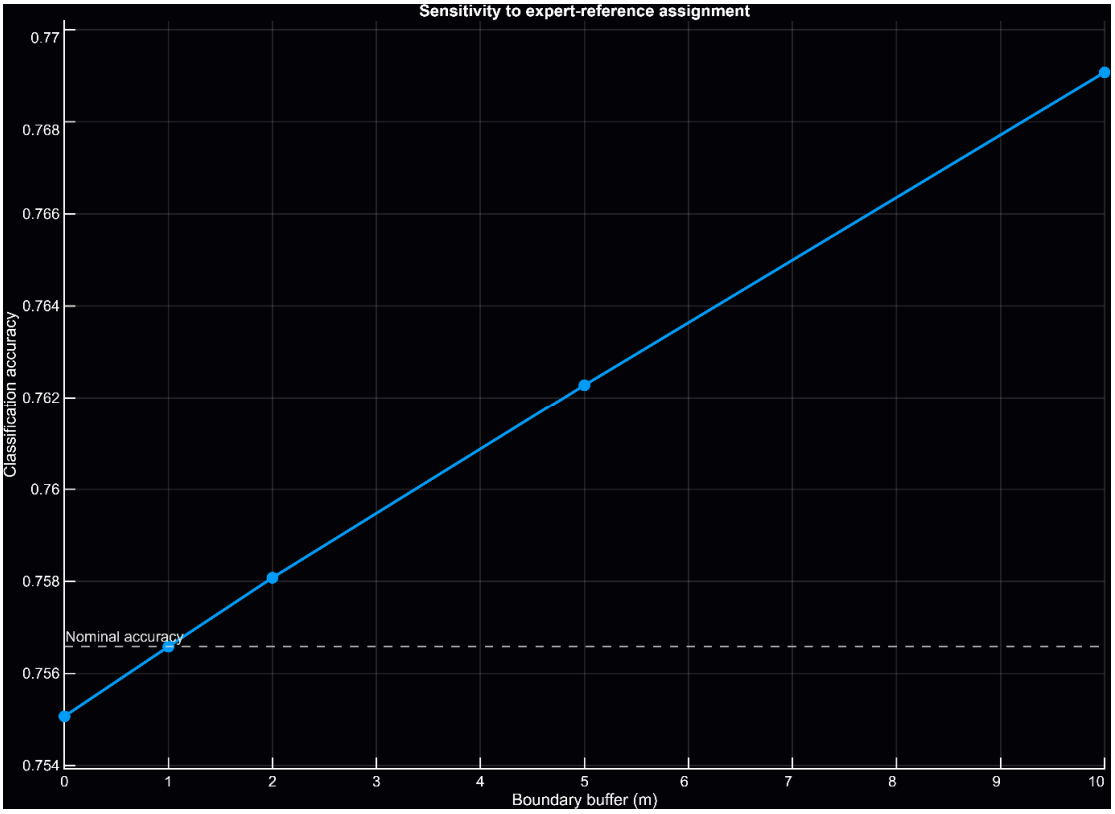

Figure 8. Sensitivity of expert-reference assignment to spatial matching criteria. Classification accuracy increased modestly with exclusion-buffer distance around expert-zone boundaries, from approximately 0.755 with no buffer to 0.769 at 10 m. Results were effectively identical across minimum crown–polygon overlap criteria ranging from 50% to 90%; consequently, the overlapping series are represented by a single response curve. The dashed horizontal line indicates classification accuracy under the nominal reference-assignment settings.

4.6. Ordinal Regression and Spatial Cross-Validation

Because the expert classes represented an ordered progression in damage severity, proportional-odds ordinal regression models were evaluated using spatially blocked cross-validation. Under the primary five-fold spatial cross-validation, the combined NDRE–NRCT model (M12) achieved the highest balanced accuracy among the eligible candidate models (Figure 9). In the primary five-fold k-means spatial cross-validation, M12 achieved a balanced accuracy of 0.663 and macro-F1 of 0.672, compared with 0.651 and 0.659, respectively, for the NDRE-only model (M1). Both measures were higher for M12 in each of the five folds, although the fold-specific improvements were generally modest. This modest advantage

was specific to the primary k-means spatial partitioning; however, under the more stringent directional sensitivity analysis, M12 and M1 performed similarly and no consistent benefit from adding NRCT was observed (Section Directional Spatial Validation Sensitivity).

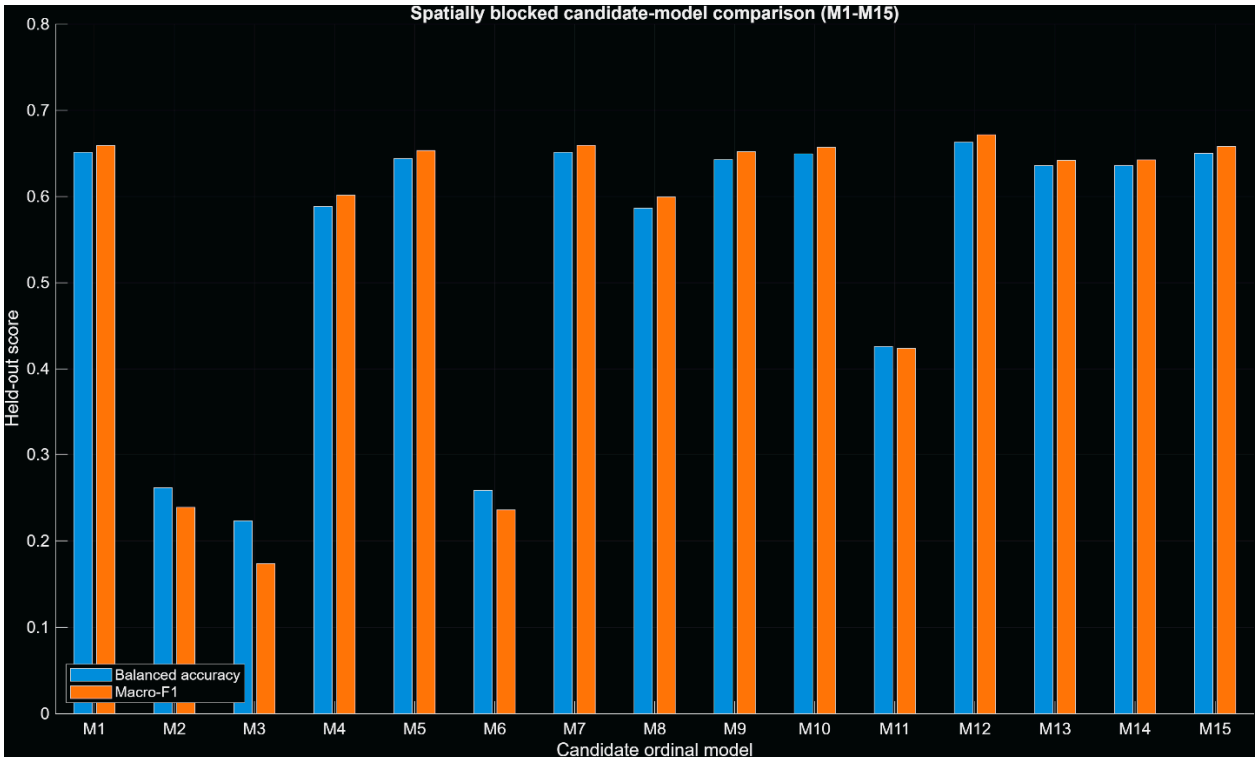

Figure 9. Comparison of candidate ordinal logistic regression models (M1–M15) using spatially cross-validated balanced accuracy and macro-F1. Models M13–M15 were added to evaluate the standalone discriminatory performance of GLI and the incremental contribution of LAI when combined additively with GLI or NDRE. All models were evaluated using identical spatial validation folds and the same 151,022 expert-referenced trees.

The incremental contribution of LiDAR-derived LAI was evaluated using additive and multiplicative formulations within the original five-fold spatial cross-validation framework. GLI alone (M13) achieved a balanced accuracy of 0.636 and macro-F1 of 0.642, compared with 0.636 and 0.643 for GLI + LAI (M14), respectively. Incorporating LAI as an independent predictor produced a negligible change in balanced accuracy (+0.0002; 95% spatial-cluster bootstrap interval: −0.0022 to 0.0027). $ICCI_G$ (M5) achieved a balanced accuracy of 0.644 and macro-F1 of 0.653, representing a modest improvement over GLI alone (+0.0082), although the corresponding bootstrap interval (−0.0219 to 0.0367) included zero. Furthermore, $ICCI_G$ exhibited higher log loss than GLI alone (0.738 versus 0.695), indicating poorer overall probabilistic predictive performance (Supplementary Table S5).

For the NDRE-based formulations, NDRE alone (M1) achieved a balanced accuracy of 0.651 and macro-F1 of 0.659, compared with 0.650 and 0.658 for NDRE + LAI (M15), respectively. The paired difference in balanced accuracy was −0.0011 (95% spatial-cluster bootstrap interval: −0.0041 to 0.0009). The multiplicative ICCI formulation (M4) performed substantially worse than NDRE alone, with balanced accuracy decreasing to 0.588 and macro-F1 to 0.602. The paired difference in balanced accuracy was −0.0629 (95% interval: −0.1113 to −0.0090). These results indicate that incorporating LAI did not consistently improve classification beyond the corresponding spectral index alone. The paired spatial-bootstrap comparisons for all additive and multiplicative formulations are summarised in Supplementary Table S6.

NDRE and NRCT exhibited moderate inverse correlation, while collinearity in the combined model remained acceptable (VIF = 2.31). Under the primary spatial partitioning, these results suggest that NRCT provided some complementary information to NDRE despite its comparatively weak standalone discriminatory performance.

Adding LAI and CD to NDRE produced little improvement. Model M7 (NDRE + LAI + crown diameter) yielded an accuracy of 0.705, balanced accuracy of 0.651, and macro-F1 of 0.659, effectively equivalent to the NDRE-only model. $ICCI_G$ alone (M5) ranked third, with an accuracy of 0.694 and balanced accuracy of 0.644, whereas ICCI alone (M4) achieved 0.642 and 0.588, respectively. Models based solely on LAI or crown diameter performed substantially worse.

The full model containing NDRE, LAI, CD, and $ICCI_G$ (M10) was excluded from model selection because of strong multicollinearity (maximum VIF = 6.03); its spatially cross-validated performance did not exceed that of the simpler NDRE-only model.

The LiDAR-derived LAI sensitivity analysis demonstrated substantial dependence on return-selection methodology. Median crown-level LAI increased from 3.803 for the original all-return estimator (E0) to 6.567 for the first-return estimator (E1) and 4.305 for the recorded-pulse proxy (E2), while the proportion of estimates exceeding LAI 10 increased from 0.34% to 31.54% and 30.37%, respectively. Pearson correlations with E0 were moderate ($r = 0.481$ for E1 and $r = 0.465$ for E2), indicating that return-selection assumptions materially influenced numerical LAI estimates (Supplementary Table S7a). However, these differences translated into only modest changes in spatially cross-validated classification performance. Replacing E0 with either alternative increased balanced accuracy for the NDRE + LAI model (M15) from 0.650 to 0.662, but the paired spatial-bootstrap confidence intervals included zero (Supplementary Table S7b). Direct incorporation of gap fraction alongside NDRE did not improve classification performance relative to NDRE alone (Supplementary Table S7c), while gap-fraction estimates varied with canopy-height threshold and return-selection convention (Supplementary Table S7d). Overall, alternative LiDAR-derived canopy-density formulations did not establish a robust improvement in expert-reference damage-severity classification. Furthermore, because independent field LAI measurements were unavailable, differences between estimators could not be interpreted as differences in physiological accuracy.

Alternative LAI estimators also affected descriptive damage-class separation. The absolute standardised difference between low and severe damage classes increased from 0.779 for E0 to 1.479 and 1.626 for E1 and E2, respectively. Nevertheless, these changes did not produce a spatially robust improvement in classification. For the additive NDRE–LAI model (M15), balanced accuracy increased from 0.650 using E0 to 0.662 using either E1 or E2, while macro-F1 increased from 0.658 to 0.670. However, paired spatial-cluster bootstrap intervals for these improvements included zero. The incremental balanced-accuracy improvement over NDRE alone was approximately 0.010 for both alternative estimators (95% spatial-bootstrap interval: −0.009 to 0.025). Thus, the alternative estimators did not establish a robust additional contribution from LAI to spatially validated classification.

The supplementary gap-fraction analysis provided a further comparison of the underlying structural information. The original all-return gap fraction was mathematically complementary to the quantity previously termed canopy cover, which was therefore identified as a canopy return fraction rather than an independent geometric cover measure. Gap fraction and LAI were deterministically related at the raster-cell level, although nonlinear transformation and subsequent crown averaging produced different crown-level summaries. The original LAI estimator showed greater absolute low–severe class separation than direct gap fraction (standardised differences of 0.779 and 0.626, respectively). In spatial cross-validation, balanced accuracy was 0.651 for NDRE alone, 0.650 for NDRE combined with

original LAI, and 0.646 for NDRE combined with gap fraction. The latter represented a balanced-accuracy difference of −0.005 relative to NDRE alone (95% spatial-bootstrap interval approximately −0.010 to 0.000). Direct gap fraction therefore provided no improvement over the original LAI formulation or the spectral baseline.

In the selected NDRE–NRCT proportional-odds model (M12), standardised NDRE had a fitted coefficient of 3.015, corresponding to an odds ratio of 20.38 per one-standard-deviation increase for membership in a lower rather than higher damage category. Standardised NRCT had a coefficient of −0.476, corresponding to an odds ratio of 0.621. Model-based 95% confidence intervals for the standardised coefficients were 2.986–3.043 for NDRE and −0.499 to −0.454 for NRCT. Spatial-cluster bootstrap resampling produced substantially wider percentile intervals of 1.977–3.915 for NDRE and −1.007 to 0.133 for NRCT. The NDRE coefficient therefore remained consistently positive under both uncertainty approaches, whereas the spatial-bootstrap interval for NRCT included zero. NRCT consequently provided modest incremental predictive information when combined with NDRE under the primary *k*-means spatial cross-validation, as evidenced by improvement in all five primary spatial folds, but the present data provide weaker evidence for a spatially robust independent thermal association after accounting for NDRE.

The conventional Brant test strongly rejected the proportional-odds assumption for M12 ($\chi^2(2) = 2979.330$, $p < 10^{-300}$, Supplementary Table S8a). Significant departures from parallel slopes were identified for both NDRE ($\chi^2(1) = 602.288$, $p = 5.32 \times 10^{-133}$) and NRCT ($\chi^2(1) = 2906.884$, $p < 10^{-300}$). Boundary-specific standardised cumulative-logit coefficients nevertheless showed that NDRE retained the same direction of association at both ordinal boundaries, increasing from 2.579 for low versus medium-or-severe damage to 3.266 for low-or-medium versus severe damage. In contrast, the NRCT coefficient changed from −1.121 at the lower boundary to 0.024 at the upper boundary, indicating that its conditional association was concentrated primarily at the lower damage transition and was negligible at the severe-damage transition. These results indicate substantial departures from the common-slope assumption under conventional tree-level inference, although the statistical significance of the tests should be interpreted cautiously because neighbouring trees are spatially correlated (Supplementary Table S8a).

An exploratory spatial-cluster sensitivity analysis was also undertaken to assess the influence of spatial dependence on the proportional-odds diagnostic. Using a sandwich covariance estimator based on ten spatial clusters, the global Wald statistic decreased to $\chi^2(2) = 3.357$ ($p = 0.187$), with corresponding predictor-specific statistics of $\chi^2(1) = 0.769$ ($p = 0.381$) for NDRE and $\chi^2(1) = 2.533$ ($p = 0.111$) for NRCT. However, inference based on only ten large, spatially heterogeneous clusters is potentially unstable and insufficient to establish that the proportional-odds assumption holds. These results therefore indicate sensitivity of the conventional significance tests to spatial dependence rather than providing definitive evidence for or against proportional odds.

The proportional-odds assumption was strongly rejected for M12 (Brant test: $\chi^2(2) = 2979.330$, $p < 10^{-300}$). Predictor-specific tests also indicated departures from proportional odds for both NDRE ($\chi^2 = 602.288$) and NRCT ($\chi^2 = 2906.884$). However, NRCT exhibited the larger qualitative difference between ordinal boundaries: its diagnostic cumulative-logit coefficient changed from −1.121 for the low versus medium/severe boundary to 0.024 for the low/medium versus severe boundary, whereas the corresponding NDRE coefficients were 2.579 and 3.266 (Supplementary Table S8b). These results indicate that the associations between the predictors and expert-assigned damage severity were not constant across the ordinal thresholds.

Relaxing the proportional-odds assumption produced modest changes in spatially validated classification performance (Table 3). The original M12 model achieved a balanced

accuracy of 0.663 and macro-F1 of 0.672. PPO-1, which allowed NRCT to vary between ordinal boundaries, increased these values to 0.672 and 0.680, respectively. PPO-2 provided no meaningful additional improvement. The corresponding coefficient estimates for the proportional-odds, partial proportional-odds, and multinomial models are presented in Supplementary Table S8c. Multinomial logistic regression produced the highest point estimates of balanced accuracy (0.687) and macro-F1 (0.692) but also had the highest log loss (0.710), compared with 0.668 for the original model. Paired spatial-cluster bootstrap intervals for the classification improvements included zero for all alternative specifications (Supplementary Table S8d). PPO-1 produced modest improvements in log loss and Brier score, although their bootstrap intervals also included zero. Multinomial regression significantly worsened log loss relative to the original model (difference = 0.0416; 95% interval: 0.0005–0.0829). Fold-level performance for the four model specifications is reported in Supplementary Table S8e. Additional class-specific discrimination results, including precision, recall, and specificity, are provided in Supplementary Table S8f. Class-specific probability-calibration statistics for the alternative model specifications are reported in Supplementary Table S8g.

Table 3. Sensitivity of spatially cross-validated M12 classification performance to the proportional-odds assumption. PPO-1 allows NRCT coefficients to vary across ordinal boundaries, whereas PPO-2 allows both NDRE and NRCT coefficients to vary. Performance metrics were calculated from pooled out-of-fold predictions using identical spatial folds. Lower log loss and Brier scores indicate better probability predictions. Bold indicates best score.

Model	Balanced Accuracy	Macro-F1	Log Loss	Brier Score
M12: Proportional odds	0.663	0.672	0.668	0.393
PPO-1: NRCT relaxed	0.672	0.680	**0.662**	**0.388**
PPO-2: Both predictors relaxed	0.672	0.681	0.681	0.391
Multinomial logistic	**0.687**	**0.692**	0.710	0.394

Relative to the original M12 predictions, PPO-1 changed the assigned damage class for 3623 trees (2.40%), while PPO-2 changed 3520 (2.33%) and multinomial regression changed 14,463 (9.58%). All changes occurred between adjacent severity classes, with no direct low-to-severe or severe-to-low transitions. For PPO-1, prediction changes were more frequent near expert-delineated damage-zone boundaries, with class-change rates of 3.38% and 0.72% in the nearest and most distant boundary-distance quintiles, respectively. These findings indicate that the principal spatial classification patterns were relatively stable under the PPO specifications, although the multinomial model produced more extensive changes.

The sensitivity of PPO-1 to the constraint domain was also examined. Restricting the constraints to the observed predictor support permitted greater variation in the boundary-specific NRCT coefficients and increased spatially cross-validated balanced accuracy from 0.672 to 0.675. However, the improvement relative to the original proportional-odds M12 remained uncertain under paired spatial-cluster bootstrap resampling (Δ balanced accuracy = +0.012; 95% interval: −0.002 to 0.021). Thus, the principal conclusion regarding the absence of a spatially robust classification improvement was unchanged.

The original NRCT raster and crown-level thermal measurements were reproduced exactly from the retained source data. Alternative non-clipped normalisation bounds, including canopy-restricted extrema and percentile-based limits, produced affine transformations of the original NRCT values. Consequently, crown-level rankings, ROC-AUC, and refitted ordinal model predictions remained unchanged within numerical precision (maximum predicted-probability difference = 6.35×10^{-10}).

Canopy-restricted extraction produced similarly small differences. Relative to the original full-crown extraction, LiDAR-derived masks of LAI > 0 and LAI > 0.5 yielded crown-level NRCT correlations of 0.999927 and 0.999321, respectively. On the common evaluation population of 148,730 trees, spatially cross-validated balanced accuracy for M12 was 0.6675 using the original extraction, compared with 0.6675 and 0.6679 for the two canopy-restricted alternatives. The corresponding improvements over NDRE alone (M1) were 0.0133, 0.0133, and 0.0137. The difference between the LAI > 0.5 configuration and the original extraction was only 0.0004 (95% spatial-cluster bootstrap interval: −0.0002 to 0.0006). The LAI > 0.5 mask retained 98.48% of reference trees, with slightly lower retention among severely damaged crowns (97.72%) than among low (98.70%) and medium (99.49%) damage classes. Overall, the tested normalisation and canopy-restriction procedures did not materially alter the incremental classification contribution of NRCT.

Directional Spatial Validation Sensitivity

The more stringent directional spatial-validation analysis revealed substantial spatial heterogeneity in predictive performance (Figures S1–S4; Tables S1–S3). Across the five contiguous X-coordinate holdouts, the arithmetic mean of the class-represented fold balanced-accuracy values was 0.692 for the NDRE-only model (M1) and 0.688 for the NDRE–NRCT model (M12), while the corresponding arithmetic mean macro-F1 values were 0.608 and 0.606.

These aggregate values are reported only as spatial-sensitivity summaries because all five directional holdouts were class incomplete and therefore do not represent estimates of three-class predictive performance. The mean M12–M1 differences were −0.0037 for balanced accuracy and −0.0024 for macro-F1. M12 did not show a consistent advantage over M1: balanced accuracy was higher for M12 in one band, higher for M1 in one band and effectively tied in three, while macro-F1 favoured M12 in one band, M1 in two, and was effectively tied in two.

The class incompleteness arose from the strong spatial segregation of expert-reference damage severity along the west–east gradient, such that individual X-coordinate bands did not provide independent three-class validation sets (Figure S1). Performance consequently varied markedly among bands, reflecting differences in both class composition and spatial position. The directional analysis should therefore be interpreted primarily as a sensitivity test of spatial transfer rather than as an alternative estimate of overall three-class classification accuracy.

The reciprocal edge-to-edge tests further demonstrated this asymmetry. Training on the western 80% and predicting the eastern 20% produced apparent balanced accuracies of approximately 0.91 for both models; however, the eastern holdout contained only severe reference trees and consequently did not provide a valid test of three-class discrimination. In the reverse direction, training on the eastern 80% and predicting the western 20% reduced balanced accuracy to approximately 0.514–0.517 and macro-F1 to approximately 0.331 for both models. M1 and M12 performed similarly in both directional tests.

The observed degradation was not primarily attributable to extrapolation beyond the absolute predictor ranges represented during training. At most, 0.056% of validation observations fell outside the corresponding training minimum–maximum range. Greater distributional shifts were evident within the tails, however, with up to 17.2% of validation NDRE observations and 4.8% of NRCT observations falling outside the training 1st–99th percentile range (Figure S4). Overall, the directional sensitivity analysis did not support a consistent predictive advantage of adding NRCT to NDRE under stronger spatial separation.

4.7. Classification Agreement and Class-Specific Discrimination

The selected NDRE–NRCT ordinal classifier showed the greatest class-specific agreement with the expert assessment for severe damage (Figure 10). Of the 72,654 trees assigned to the severe-damage class by expert assessment, 64,054 were correctly classified as severe, corresponding to a recall of 88.2%. A further 8558 were classified as medium and only 42 as low damage. Direct confusion between the severe- and low-damage classes was therefore minimal.

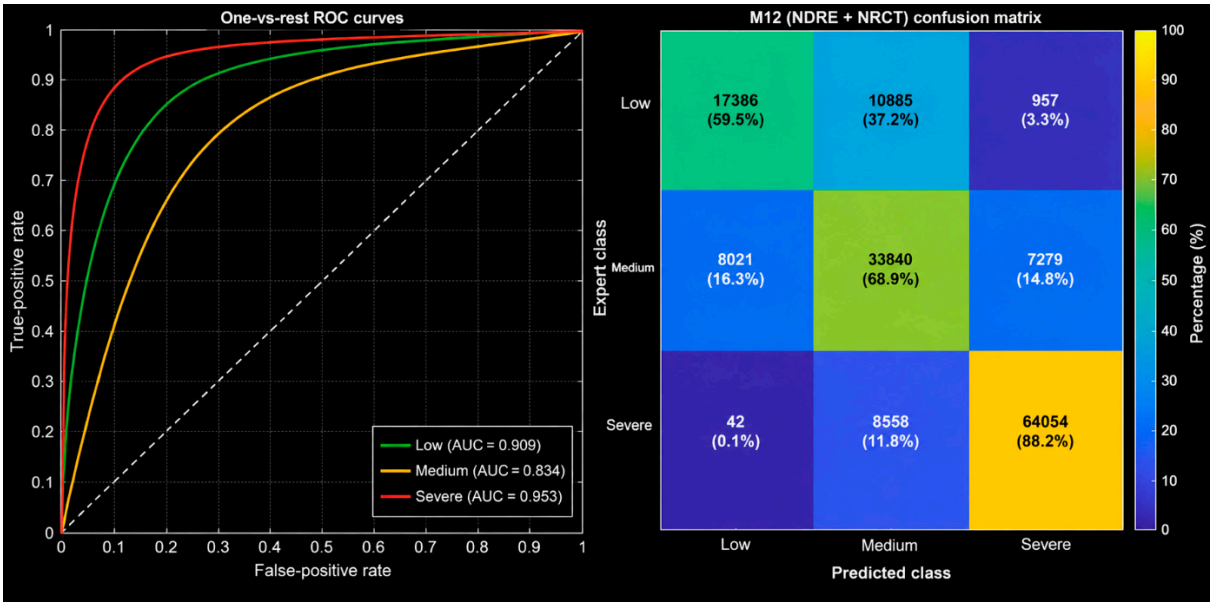

Figure 10. Validation of the selected NDRE–NRCT ordinal hail-damage classifier against the expert reference assessment: (**left**) one-versus-rest ROC curves for low, medium, and severe damage, with AUC values of 0.909, 0.834, and 0.953, respectively; and (**right**) confusion matrix comparing model-predicted and expert-assigned damage classes. Most classification errors occurred between adjacent severity classes, with little direct confusion between low and severe damage.

Of the 49,140 expert-assigned medium-damage trees, 33,840 (68.9%) were correctly classified, while 8021 were assigned to low damage and 7279 to severe damage. Of the 29,228 low-damage trees, 17,386 (59.5%) were correctly classified, 10,885 were classified as medium, and 957 as severe. Overall, classification errors occurred predominantly between adjacent severity classes: 34,743 of 35,742 errors (97.2%) involved low–medium or medium–severe confusion. Direct low–severe confusion was limited and asymmetric, comprising 3.3% of expert-assigned low trees but only 0.1% of expert-assigned severe trees.

One-versus-rest ROC analysis showed the highest discrimination for severe damage (AUC = 0.953), followed by low damage (AUC = 0.909), with the medium class showing the lowest discrimination (AUC = 0.834).

Agreement obtained after refitting the selected model to the complete reference dataset represents correspondence with the calibration data and should not be interpreted as spatially independent predictive accuracy. Five-fold spatial cross-validation produced a balanced accuracy of 0.663 and macro-F1 score of 0.672 for the selected NDRE–NRCT model. These spatially held-out results therefore provide the principal and more conservative estimate of predictive performance within the analysed plantation.

4.8. Spatial Pattern of Predicted Damage

Tree-level probability maps reproduced the broad spatial organisation of the expert assessment (Figure 11). High probabilities of low damage were concentrated within the

least-affected portion of the plantation, whereas high severe-damage probabilities predominated within the heavily affected area. Elevated probabilities of medium damage occurred principally within the transition between these regions and along locally irregular class boundaries.

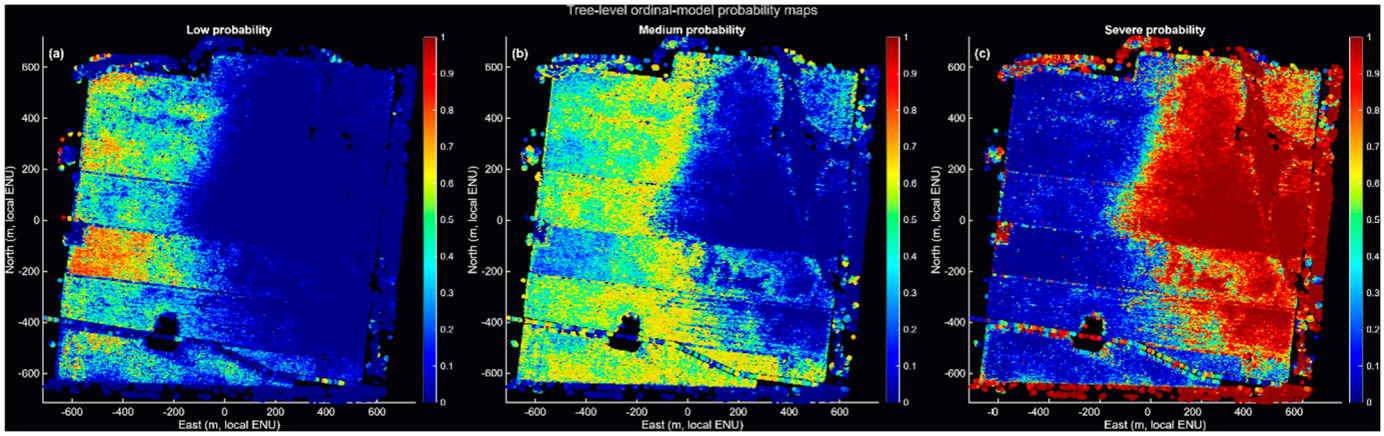

Figure 11. Spatial distribution of predicted hail-damage probabilities at individual-tree level. Predicted probabilities of (**left**) low, (**centre**) medium, and (**right**) severe hail damage for individual segmented trees. Low- and severe-damage probabilities dominate opposite portions of the plantation, while elevated medium-damage probabilities occur principally within the intervening transition zone. Coordinates are expressed in the local plantation coordinate system.

Within the interiors of the low and severe-damage zones, individual trees generally exhibited high probabilities for the corresponding class. In contrast, trees near damage-zone boundaries showed more distributed probabilities among adjacent classes, providing a spatial representation of classification uncertainty not conveyed by the discrete expert polygons.

Crown-level NDRE measurements were highly consistent between the original 1 m grid and the independently aggregated 0.25 m grid (Pearson r = 0.9936; Spearman ρ = 0.9912; mean absolute difference = 0.0108). On the common-tree evaluation population (n = 145,823), the corresponding spatially cross-validated balanced accuracies were 0.6564 and 0.6547, respectively, representing a difference of −0.0017 (95% spatial-cluster bootstrap interval: −0.0059 to 0.0020). Severe-damage AUC was similarly stable (0.9579 and 0.9584). Excluding crown-boundary observations using inward buffers of 0.25 m and 0.50 m reduced balanced accuracy to 0.6268 and 0.6148, respectively, while severe-damage AUC remained approximately 0.958–0.959. The severe-damage NDRE threshold was approximately 0.534 for the full-crown configurations but increased to 0.547–0.549 following boundary exclusion. These findings indicate that the principal NDRE–damage relationship was insensitive to the tested change in full-crown aggregation resolution, although crown-boundary exclusion affected class distributions, classification performance, and calibrated thresholds.

5. Discussion

The principal finding of this study was that expert-assessed hail severity could be reproduced quantitatively at individual-tree scale using UAV-derived observations, but the discriminatory information was dominated by spectral rather than conventional structural canopy metrics. This distinction has implications both for interpretation of hail-induced canopy responses and for the design of operational post-event assessment workflows.

5.1. Spectral and Thermal Responses to Hail Damage

NDRE provided the clearest standalone separation among expert-assigned damage classes and remained the dominant predictor of hail-damage severity. Mean NDRE declined from approximately 0.65 in the low-damage class to 0.59 under medium damage and 0.40 under severe damage, and the severe-damage transition achieved an AUC of 0.954. However, the highest spatially cross-validated performance was obtained when NRCT was included alongside NDRE.

NRCT provided a modest improvement when added to NDRE under the primary five-fold k-means spatial cross-validation, but this advantage was not maintained under the more stringent directional sensitivity analysis. M12 improved balanced accuracy and macro-F1 relative to M1 in all five primary spatial folds, whereas the contiguous X-band analysis produced essentially equivalent mean performance for the two models and no consistent fold-wise advantage for M12. The evidence for complementary predictive information from canopy temperature is therefore dependent on the spatial validation design and should be interpreted as incremental rather than robust across alternative spatial partitions.

The results indicate that the majority of the discriminatory information was contained in NDRE alone. The addition of thermal or structural predictors produced little improvement in spatially validated classification performance, suggesting that the principal benefit of multisensor integration in this application was not increased severity-classification accuracy. Rather, the complementary sensors provided additional information for individual-tree delineation, structural characterisation, and potential integration with forest inventory systems. Consequently, the value of multisensor acquisition should be assessed in relation to the intended operational outputs and associated acquisition and processing costs, rather than classification accuracy alone.

The boundary-specific analysis further clarified the different roles of the two predictors. NDRE retained a strong positive association at both ordinal transitions, indicating that its relationship with damage severity was directionally consistent across the full severity gradient, although its magnitude differed between transitions. In contrast, the conditional NRCT association was concentrated primarily at the low versus medium-or-severe transition and was negligible at the low-or-medium versus severe transition. Thus, the thermal contribution did not represent a uniform response across damage severity but was associated principally with the transition between relatively lightly damaged trees and those exhibiting greater crown damage.

This interpretation is consistent with the coefficient uncertainty. Although the fitted NRCT coefficient was negative and its conventional model-based confidence interval excluded zero, its spatial-cluster bootstrap interval included zero. Together, the bootstrap and directional-validation results indicate substantially weaker evidence for an independent and spatially transferable thermal contribution than for NDRE. NRCT may capture local variation in canopy condition not represented completely by the red-edge response, but the present analysis does not demonstrate that this additional information transfers consistently across the strong spatial damage gradient within the plantation.

The strong NDRE response is consistent with the expected effects of severe hail on tree crowns. Hail can remove needles and small branches and damage remaining needle foliage, thereby reducing the amount and condition of photosynthetically active canopy material. Because red-edge reflectance is particularly sensitive to chlorophyll concentration and vegetation condition, these effects would be expected to reduce NDRE. The relatively small difference between the low- and medium-damage classes compared with the much larger reduction under severe damage suggests a non-linear spectral response once substantial crown deterioration occurs. This interpretation is also consistent with the greater

classification confusion between low and medium damage than between either of these classes and severe damage.

The additional model comparisons further clarify the limited contribution of LiDAR-derived LAI to severity classification. GLI alone achieved a spatially cross-validated balanced accuracy of 0.636, compared with 0.644 for $ICCI_G$, although this difference was uncertain under spatial-cluster bootstrap resampling (95% confidence interval: −0.022 to 0.037). Incorporating LAI as a separate predictor alongside GLI produced negligible improvement. Similarly, adding LAI to NDRE did not improve classification, while the multiplicative ICCI formulation substantially reduced balanced accuracy from 0.651 to 0.588. These findings suggest that the spectral indices captured most of the information relevant to the expert-assessed damage classes and that the tested spectral–structural combinations offered limited additional discriminatory value. Nevertheless, LiDAR remains valuable for individual-tree detection, crown delineation, and structural characterisation, supporting the spatial attribution and operational interpretation of damage assessments.

The LiDAR sensitivity analyses highlight an important distinction between the physical interpretation of canopy-structure measurements and their utility for hail-damage classification. The baseline LiDAR-derived LAI estimator, calculated using all recorded returns (E0), was based on the proportion of returns classified as canopy and therefore represented an effective structural interception or plant-area proxy rather than an independently validated measure of foliage area. Its magnitude was strongly dependent on the treatment of multiple returns, with first return and recorded-pulse formulations producing substantially different LAI distributions. These differences were also associated with canopy height, return multiplicity and LiDAR sampling density, indicating that acquisition and processing characteristics influence the resulting structural measurements.

Although alternative LAI estimators produced greater descriptive separation between low and severe damage classes, their apparent improvements in classification were not robust under spatial-cluster bootstrap uncertainty. Direct gap fraction, which avoids the logarithmic transformation and assumed extinction coefficient used in LAI estimation, also failed to improve spatially validated classification relative to NDRE alone. The quantity previously designated canopy cover was found to be an exact complement of gap fraction and did not represent an independent measure of projected canopy cover. These findings reinforce the conclusion that the evaluated LiDAR-derived canopy-density measures provided limited additional information beyond the spectral indices for predicting expert-assigned hail-damage severity.

The results should not be interpreted as demonstrating that canopy structural changes are unimportant following hail damage. Needle loss, branch damage, and changes in crown openness may affect LiDAR measurements differently, while the present return-based estimators cannot distinguish foliage from woody components. Furthermore, the recorded-pulse analysis was constrained by incomplete pulse information, and independent field measurements of LAI or needle loss were unavailable. Consequently, the analyses establish sensitivity to measurement methodology and limited incremental classification performance, rather than the physiological accuracy of any particular estimator. Future investigations incorporating emitted-pulse information and independent foliage or canopy-gap measurements would help determine whether more physically representative structural metrics offer additional diagnostic value.

The modest predictive contribution of NRCT may reflect the different aspects of canopy condition represented by spectral and thermal observations. Whereas NDRE is primarily sensitive to red-edge reflectance associated with needle amount, chlorophyll status, and overall crown condition, NRCT represents spatial variation in relative canopy surface temperature, which can be influenced by transpiration, canopy exposure, and local energy

balance. The negative fitted NRCT coefficient is physiologically consistent with relatively warmer canopy conditions being associated with greater expert-assessed damage severity after accounting for NDRE. The moderate inverse correlation between the two predictors indicates that their responses were related but not redundant, suggesting that thermal observations may capture aspects of post-hail canopy response not completely represented by NDRE. However, canopy temperature is also influenced by environmental and spatial variability unrelated to hail damage, which may partly explain the weaker and less consistent contribution of NRCT.

The strong performance of NDRE should not, however, be interpreted as evidence that NDRE measures hail damage directly. Spectral indices respond to vegetation condition rather than to the causal disturbance itself. Changes in needle density, chlorophyll concentration, illumination, crown geometry, and background reflectance can all influence measured values. In the present plantation, the close spatial correspondence between declining NDRE and independently mapped hail severity provides strong evidence that hail was the principal source of the observed gradient, but application to other sites would require appropriate reference observations and recalibration.

The fitted common-slope effects in the selected NDRE–NRCT model further illustrate the difference in the overall magnitude and robustness of the two predictors. Under the proportional-odds parameterisation, a one-standard-deviation increase in NDRE corresponded to an estimated cumulative odds ratio of 20.4, whereas the corresponding odds ratio for NRCT was 0.621. Because the proportional-odds assumption was rejected, however, these values should be interpreted as model-level summary effects rather than as transition-invariant effect sizes. Boundary-specific coefficients showed that the positive NDRE association persisted at both damage transitions (2.579 and 3.266), whereas the NRCT association differed substantially between them (−1.121 and 0.024). Model-based 95% confidence intervals for the common-slope coefficients were 2.986–3.043 for NDRE and from −0.499 to −0.454 for NRCT, whereas spatial-cluster bootstrap intervals were substantially wider at 1.977–3.915 and from −1.007 to 0.133, respectively. Taken together, the boundary-specific and spatial-bootstrap analyses indicate a strong and directionally consistent association between NDRE and damage severity, whereas the independent contribution of NRCT was less robust and strongly dependent on the ordinal transition. The magnitude of the NDRE odds ratio should consequently be interpreted as evidence of strong class separation rather than as a directly transferable biological response parameter, because it remains conditional on the predictor distributions, scaling, and class definitions used in this plantation.

The spatial-resolution sensitivity analysis provides additional evidence that the observed relationship between NDRE and expert-assessed hail damage was not an artefact of the original 1 m analysis grid. Independently aggregating calibrated spectral observations to 0.25 m produced nearly identical crown-level measurements and spatially validated classification performance. Restricting extraction to crown interiors did not eliminate severe-damage discrimination but reduced three-class classification performance and altered the empirical thresholds. These effects may reflect changes in the representation of foliage, canopy gaps, exposed branches, understory, and background surfaces within damaged crowns. Importantly, inward buffering removes crown-edge observations but does not eliminate background contributions visible through gaps within the crown. The resulting NDRE measurements should therefore be interpreted as indicators of overall crown condition rather than isolated measurements of foliage physiology.

The strong rejection of the proportional-odds assumption for M12 indicates that the relationships between NDRE, NRCT, and expert-assigned damage severity differ across ordinal boundaries. In particular, the association of NRCT with damage severity was

substantially different between the low versus medium/severe and low/medium versus severe comparisons. Consequently, the common coefficients estimated by the proportional-odds model should be interpreted as constrained summaries rather than uniform effects across all severity transitions.

Nevertheless, the additional modelling demonstrated that relaxing proportional odds did not produce a spatially robust improvement in classification. The partial proportional-odds model allowing NRCT to vary between boundaries produced modest improvements in balanced accuracy, macro-F1, and probability scores, but the paired spatial-bootstrap intervals included zero. Allowing NDRE to vary provided no further benefit. Although multinomial regression increased the point estimates of balanced accuracy and macro-F1, these gains were not spatially robust, and its predicted probabilities exhibited significantly worse log loss. The original M12 model was therefore retained as the principal, more parsimonious classification model, with the alternative specifications providing evidence that the main predictive conclusions were not materially dependent on the proportional-odds constraint.

Although the conventional Brant test strongly rejected the proportional-odds assumption, its tree-level covariance formulation does not account for spatial dependence among neighbouring observations. An exploratory spatial-cluster sensitivity analysis produced substantially weaker evidence against parallel slopes, although the limited number of clusters precludes definitive inference. Consequently, the conventional test results should not be interpreted in isolation as evidence that the proportional-odds model is unsuitable for operational classification. The spatially cross-validated comparisons provide a more directly relevant assessment of predictive consequences: relaxing proportional odds produced modest improvements in some classification metrics, but these were not spatially robust. M12 was therefore retained as the principal parsimonious model, with its common coefficients interpreted as constrained summaries of associations rather than evidence that predictor effects are identical across damage-class boundaries.

These findings do not establish that the proportional-odds assumption is valid. Rather, they indicate that the consequences of its violation for the principal spatial classification results were limited under the evaluation framework employed. For applications requiring accurate class probabilities or interpretation of individual predictor effects at specific damage-severity boundaries, a partial proportional-odds formulation may be more appropriate. The alternative-model results should also be interpreted in the context of the expert-derived reference labels, which represent polygon-assigned damage categories rather than independently verified observations of physical damage to individual trees.

5.2. Structural Response to Hail Damage

The LiDAR-derived structural variables generally changed in the direction expected with increasing hail damage but were substantially weaker individual predictors than NDRE. Mean LAI declined from approximately 4.36 in the low-damage class to 3.38 in the severe class, while mean crown diameter declined from 3.28 to 3.01 m. These reductions are physically consistent with the removal of needles and fine branches during a severe hail event.

The relatively modest discriminatory performance of LAI is noteworthy because LAI was expected to provide a direct structural measure of needle loss. Several factors may explain this result. First, LiDAR-derived LAI is an indirect estimate based on canopy penetration and gap fraction rather than a direct measurement of leaf area. It is consequently sensitive to point density, scan geometry, canopy clumping, and the assumptions underlying the Beer–Lambert relationship. The pronounced upper tails in the LAI distributions also indicate that a relatively small number of cells produced unusually high estimates. These

factors increase within-class variability and reduce the effectiveness of LAI as a stand-alone damage classifier.

Second, the expert-reference classes represented visually assessed crown-damage severity rather than quantitative measurements of foliage loss or individual structural attributes. Trees assigned to the same expert class could therefore retain different amounts and spatial distributions of crown material while presenting a similar overall level of visible crown damage. Conversely, LiDAR-derived structural characteristics vary naturally among trees because of competition, microsite conditions and pre-event crown development. Without pre-hail observations, this pre-existing structural variability cannot be separated completely from storm-related change using a single post-event LiDAR acquisition.

Crown diameter showed even greater overlap among classes. This is unsurprising because hail is unlikely to reduce the maximum horizontal crown extent uniformly. Loss of needle and small branches can produce substantial reductions in canopy density without producing an equivalent reduction in equivalent crown diameter, particularly when some peripheral branches remain. Crown diameter is also strongly influenced by pre-disturbance tree size and competition. Consequently, its relatively poor performance as an independent classifier does not imply that structural crown damage was absent; rather, a single two-dimensional diameter measure appears insufficient to capture the complexity of the structural response.

These findings support the use of more detailed three-dimensional structural descriptors in future work. Crown volume, vertical foliage profiles, upper-canopy density, crown asymmetry, and voxel-based measures of canopy occupancy may be more sensitive to hail-induced structural changes than crown diameter alone.

5.3. Value of Integrating Spectral and Structural Information

The integrated canopy indices reproduced the expected progression from low through medium to severe damage. Both ICCI and $ICCI_G$ declined substantially with increasing expert-assessed severity, demonstrating that combining vegetation condition with LiDAR-derived effective canopy foliage density provides a physically interpretable measure of post-hail canopy condition.

An important result was that $ICCI_G$ performed better than ICCI in the direct comparison of the two integrated indices. $ICCI_G$ achieved an overall accuracy of approximately 0.742, balanced accuracy of 0.691, and weighted κ of 0.750, compared with 0.702, 0.650, and 0.707, respectively, for ICCI. This should not be interpreted as evidence that the RGB GLI is intrinsically superior to NDRE as a vegetation index. More generally, neither integrated index should be interpreted as a mechanistic measure of hail injury. They are empirical composites designed to combine complementary information on canopy density and spectral condition, and their value therefore depends on demonstrated discriminatory performance rather than on the multiplication itself having a unique biological interpretation. Indeed, NDRE considered independently was the strongest-performing metric in the study. Rather, the result suggests that the specific combination of GLI and LAI used in $ICCI_G$ provided a robust representation of the expert damage gradient.

There are several plausible explanations for this result. ICCI requires fusion of independently acquired multispectral and LiDAR datasets, and any residual registration error is propagated into the cell-by-cell product. In contrast, $ICCI_G$ can potentially be derived from RGB information associated with the LiDAR acquisition together with LAI derived from the same point-cloud dataset. Reduction of inter-sensor registration uncertainty may therefore partly compensate for the lower spectral sophistication of the red–green index. In addition, severe hail produced pronounced visible changes in canopy greenness, meaning that GLI may contain sufficient information to distinguish strongly damaged crowns

when combined with a measure of canopy density. These explanations remain hypotheses and should be tested directly using controlled comparisons of registration error and sensor combinations.

Operationally, these results distinguish maximum discriminatory performance from workflow simplicity. NDRE remains the preferred individual metric at this site, whereas $ICCI_G$ may provide a useful alternative where avoiding the acquisition and registration of separate multispectral and LiDAR datasets is advantageous. $ICCI_G$ should therefore be regarded as an operational alternative rather than a general replacement for red-edge observations.

The reduction in NDRE associated with severe hail damage should not necessarily be interpreted solely as a decline in the physiological condition of surviving foliage. Hail-induced defoliation and crown fragmentation may increase the contribution of woody material, understory vegetation and exposed ground to the measured crown-level reflectance. NDRE therefore potentially represents a combined response to foliage condition and canopy openness. This interpretation is consistent with its use as an operational indicator of crown damage but limits the extent to which the observed spectral changes can be attributed to physiological stress independently of structural foliage loss.

5.4. Empirical Thresholds and Their Robustness

A major objective of the calibration analysis was to replace visually selected damage thresholds with empirically derived values linked to the expert reference assessment. The resulting NDRE thresholds of approximately 0.606 and 0.534 provided quantitative transitions between low, medium, and severe damage, while corresponding threshold pairs were also obtained for ICCI and $ICCI_G$. Spatial-block bootstrap analysis showed relatively concentrated threshold distributions, particularly for NDRE, providing evidence that the estimated thresholds were reasonably stable to spatial resampling and were not dependent on any single spatial subset of the plantation.

The sensitivity analysis provides an important qualification, however. Classification performance was not always maximised at the nominal empirically selected thresholds. NDRE accuracy increased when the thresholds were perturbed upwards by approximately 5%, while ICCI and $ICCI_G$ exhibited broader response plateaus. This reinforces the need to treat the derived values as statistical decision thresholds rather than fixed biological boundaries. Accordingly, the bootstrap intervals quantify uncertainty in threshold estimation within the spatial structure of the analysed plantation; they should not be interpreted as transferability intervals for application to independent plantations or hail events. Hail damage occurs along a continuum, whereas the expert assessment imposes three discrete management classes onto that continuum. A range of threshold values can therefore produce similar classification outcomes.

$ICCI_G$ was particularly insensitive to moderate positive threshold perturbations, suggesting operational classifications based on this metric may be comparatively robust to modest recalibration errors. Such stability is valuable for applied forest management, where sensor conditions, illumination, stand age, and seasonal effects may prevent an exact threshold derived at one site from being reproduced elsewhere.

The weak sensitivity to crown-overlap criteria and expert-boundary exclusion buffers is also encouraging. Reasonable changes in the spatial rules used to associate crowns with expert polygons produced little change in overall performance, suggesting that the main relationships were not artefacts of a particular overlap criterion. Nevertheless, this does not remove uncertainty associated with the expert polygons themselves.

5.5. Performance of the Ordinal Damage Model

Spatially blocked cross-validation produced a balanced accuracy of approximately 0.66 for the best-performing candidate model. This represents useful predictive skill, but it is substantially more conservative than the apparent agreement obtained when the fitted classifier was applied to the complete reference dataset. The spatially blocked result should therefore be regarded as the more realistic indication of expected model performance when transferred to an independent portion of the plantation.

The directional sensitivity analysis also illustrates an important limitation of validation within a single spatially coherent disturbance event. Expert-assessed severity was strongly segregated along the west–east damage gradient, resulting in class-incomplete validation sets when geographically contiguous directional regions were withheld. Consequently, increasing spatial separation does not necessarily produce a stronger estimate of three-class predictive performance: beyond a certain degree of separation, spatial independence becomes confounded with differences in class composition. The primary k-means spatial cross-validation therefore remains the more appropriate basis for comparing candidate models within this plantation, while the directional analysis provides a deliberately stringent test of sensitivity to spatial partitioning and transfer along the damage gradient.

The combined NDRE–NRCT model ranked highest under spatially blocked cross-validation, although its improvement over the simpler NDRE-only model was modest. The addition of NRCT increased balanced accuracy from 0.651 to 0.663 and macro-F1 from 0.659 to 0.672, whereas addition of LAI and crown diameter to NDRE produced essentially no improvement. These results indicate that most of the discriminatory information required to reproduce the expert classes was contained in NDRE, with NRCT contributing a smaller amount of complementary thermal information.

The full candidate model containing NDRE, LAI, crown diameter, and $ICCI_G$ achieved favourable in-sample information criteria but exhibited substantial collinearity and did not improve spatially held-out performance. Excluding this model from selection was therefore appropriate. Because ICCI and $ICCI_G$ are constructed partly from variables also entered separately into candidate models, collinearity is an inherent concern in interpreting multivariable coefficients. Model comparison based on spatially independent prediction is consequently more informative than selecting models according to in-sample goodness-of-fit alone.

Severe damage was classified most successfully, while errors occurred predominantly between adjacent severity classes. This pattern is examined further for the transitional medium-damage class below.

5.6. Interpretation of the Medium-Damage Class

The medium class was inherently the most difficult category to define and predict. Its spectral and structural distributions overlapped with both low and severe damage, and its one-versus-rest AUC was lower than that of either extreme class. This result likely reflects both biological continuity and the nature of the expert reference assessment.

Trees near the transition between low and severe impact can exhibit combinations of structural retention and physiological stress that do not fit neatly within one category. The probability maps illustrate this behaviour clearly: high medium-damage probabilities were concentrated within transition zones rather than forming a completely discrete spatial population. Such probabilistic information may be more useful operationally than forcing every tree into a hard class.

The expert definitions themselves should also be considered when interpreting model performance. The mapped categories represent operational expert assessments of crown-damage severity rather than quantitative measurements of damage for every individual

tree. The remote-sensing models are consequently being evaluated against an informed but spatially generalised representation of a continuous disturbance process. Agreement should therefore be interpreted as the ability to reproduce and spatially refine the expert operational classification, rather than validation against error-free biological ground truth.

5.7. Operational Implications

The results demonstrate several potential advantages of the proposed workflow for operational plantation management. First, the method converts broad expert-defined damage zones into quantitative individual-tree predictions, allowing damage to be represented as tree counts, class probabilities, and spatial distributions, with the potential for subsequent integration with estimates of affected timber volume. This provides a more detailed quantitative basis for salvage planning and inventory adjustment than plantation-level visual assessment alone.

Second, the probability outputs allow classification uncertainty to be mapped explicitly. Trees with high probabilities of severe damage can be prioritised for field inspection or consideration in salvage planning, whereas transition areas characterised by ambiguous class probabilities can be targeted for additional assessment. Such spatial prioritisation may be more efficient than applying uniform field-sampling effort across an entire affected plantation and could assist in identifying priority areas from which detailed salvage assessment or operations might commence.

Third, the strong performance of NDRE indicates that relatively straightforward multispectral mapping may provide an effective rapid-assessment option where detailed structural information is not required. Conversely, $ICCI_G$ provides a potentially attractive operational pathway where RGB information and point-cloud-derived structural metrics can be obtained from the same LiDAR acquisition. If its performance is confirmed across independent sites and hail events, this approach could reduce the need to acquire, register and process a separate multispectral dataset.

The most appropriate operational product is therefore likely to depend on both the management objective and stand development stage, particularly the timber products potentially available for recovery. NDRE appears preferable where rapid discrimination of damage condition is the primary objective and may be particularly suitable for younger, unthinned stands with relatively simple product structures. Integration with LiDAR becomes increasingly valuable where individual-tree inventories, structural attributes, and estimates of recoverable or affected timber are required, as may occur in older or previously thinned stands containing multiple potential timber products. These age- and product-dependent applications were not evaluated directly in the present study and should therefore be regarded as potential operational use cases rather than demonstrated differences in performance among stand types.

5.8. Limitations

Several limitations should be considered when interpreting these results. Most importantly, the empirical calibration was derived from a single hail event, and the present expert-reference analysis is restricted to the *P. radiata* plantation. Although the analysed plantation included variation in stand age, this within-site variation does not establish transferability to substantially different plantation conditions. The thresholds and fitted relationships should therefore be regarded as site-calibrated results. Their applicability to eucalyptus or to plantations with substantially different stocking, canopy structure, or developmental characteristics, and to observations acquired in different seasons, environmental conditions, or using different sensor/acquisition configurations, remains to be established.

The pronounced spatial segregation of expert-reference severity also limits the extent to which within-plantation validation can approximate transfer to an independent hail event. All five directional holdouts contained only a subset of the three damage classes, and predictive performance varied substantially among spatial regions. Although predictor values generally remained within the ranges represented during training, shifts in their distributions were evident between regions. Independent validation across additional hail-affected plantations containing the full range of damage severity is therefore required to distinguish model transferability from spatial patterns specific to the present event.

The expert damage polygons provide an operationally relevant reference dataset but are not error-free individual-tree ground truth. They represent spatially generalised expert interpretation of crown-damage severity, and individual trees within each zone may differ in the nature and magnitude of their actual damage. This within-zone heterogeneity is likely to contribute particularly to uncertainty between the low and medium classes. Future validation using independent tree-level measurements of crown condition and specific damage attributes would provide a stronger basis for evaluating biological rather than expert-classification accuracy.

A further limitation is the absence of pre-hail UAV observations. Differences among trees are therefore inferred from spatial comparisons among post-event damage classes rather than from direct before-and-after measurements. Pre-event crown dimensions, LAI, and spectral condition may have varied across the plantation independently of the storm. Multi-temporal observations would allow actual change in individual trees to be quantified and would substantially strengthen causal attribution.

LiDAR-derived LAI also contains methodological uncertainty arising from canopy penetration, scan geometry, point density, and assumptions regarding needle distribution. Similarly, the integrated ICCI product depends on accurate spatial registration between multispectral and LiDAR observations. The unexpectedly strong performance of $ICCI_G$ highlights the possibility that co-registration uncertainty may influence the relative performance of the integrated indices and warrants further investigation.

The proportional-odds assumption was formally assessed and was strongly rejected for M12. Boundary-specific analysis nevertheless showed that NDRE retained a strong positive association at both ordinal boundaries, whereas the conditional NRCT association differed substantially between transitions and was negligible at the low-or-medium versus severe boundary. The common-slope coefficients should therefore be interpreted as model-level summary effects rather than as transition-invariant effect sizes.

Finally, the large number of individual trees should not be interpreted as an equivalent number of statistically independent observations. Neighbouring trees experienced related storm conditions and share stand and environmental characteristics, such that conventional model-based uncertainty estimates are likely to under-represent spatial uncertainty. For this reason, spatially blocked cross-validation and spatial-cluster bootstrap estimates were treated as the principal evidence for predictive performance and coefficient uncertainty, respectively. These procedures reduce, but cannot eliminate, the limitation arising from a single spatially coherent hail event; replication across independent plantations and storm events remains necessary.

5.9. Future Work

Future work should prioritise validation across independent hail events, plantation developmental stages, and species. Application of the workflow to the separately acquired eucalyptus dataset will provide a particularly informative test because broadleaf eucalyptus differs substantially from *P. radiata* in crown architecture, foliage characteristics, stocking, and exposure of background vegetation and soil, all of which may alter spectral and Li-

DAR responses. The eucalyptus dataset therefore provides an opportunity to evaluate the extent to which the present relationships are species specific and whether recalibration of individual metrics, empirical thresholds or model coefficients is required.

Tree-level field measurements should also be incorporated into future post-event surveys. Quantitative observations of crown defoliation, branch breakage, stem injury, and overall crown-damage severity would permit direct evaluation of which remotely sensed variables respond to particular manifestations of hail injury and would provide independent tree-level validation rather than relying principally on polygon-level expert reference information.

Multi-temporal UAV acquisition represents another important opportunity. Repeated surveys would enable the initial post-hail condition to be distinguished from subsequent recovery, delayed mortality, secondary physiological stress, and disease progression. Such information could determine whether trees initially classified as moderately damaged recover sufficiently to remain commercially viable or subsequently decline, thereby linking remote-sensing classifications more directly to management outcomes.

Further methodological development should examine additional three-dimensional structural variables, including crown volume, crown asymmetry, vertical foliage-density profiles, and voxel-based measures of canopy occupancy. These may capture hail-induced changes in crown structure more effectively than crown diameter or LAI alone. The contribution of registration error to ICCI performance should also be quantified directly, particularly given the comparatively strong performance of the single-acquisition $ICCI_G$ formulation.

6. Conclusions

This study developed and evaluated a UAV-based framework for quantitative post-hail damage assessment at the individual-tree scale in a commercial *Pinus radiata* plantation near Casterton, Victoria. The results demonstrate that UAV remote sensing can provide a reproducible, spatially explicit framework for post-hail plantation assessment. NDRE provided the dominant and most consistent discriminatory information, retaining the same direction of association across both ordinal damage transitions. NRCT provided more limited and context-dependent complementary information: its conditional association was concentrated primarily at the low versus medium-or-severe transition and was negligible at the transition separating severe from low or medium damage. Although adding NRCT produced a modest improvement under the primary spatial cross-validation, this advantage was not maintained under stronger directional partitioning.

In contrast, the addition of LAI and crown diameter provided little improvement over NDRE alone. Nevertheless, integration of spectral and structural information produced useful canopy-condition metrics, and the performance of the RGB–LiDAR-based $ICCI_G$ indicates potential for a simplified workflow where separate red-edge multispectral acquisition and data fusion are impractical.

Additional comparisons demonstrated that incorporating LiDAR-derived LAI, either as an independent predictor or through the multiplicative ICCI and $ICCI_G$ formulations, did not provide a robust improvement in spatially validated severity classification relative to the corresponding spectral indices alone.

Empirical calibration, spatially clustered cross-validation, cluster bootstrap resampling, and sensitivity analysis provided a framework for quantifying damage severity and uncertainty while reducing reliance on visually selected thresholds. Tree-level probability mapping further enabled spatial variation in classification confidence to be represented, particularly within transition zones between damage classes.

The workflow therefore provides a basis for rapid, spatially explicit post-hail assessment to support targeted field inspection and salvage planning, with potential for subsequent

integration with estimates of affected timber volume. The empirical thresholds and fitted models should, however, be regarded as calibrated to the analysed *P. radiata* plantation and hail event rather than as universally transferable damage relationships. Validation across independent hail events, contrasting plantation conditions, and species—including the separately acquired eucalyptus dataset—is required to establish the extent to which recalibration is necessary before broader operational application. Although the results demonstrate the potential of UAV-derived spectral and structural measurements to discriminate among expert-defined hail-damage severity zones, the reported classification performance reflects agreement with polygon-derived reference labels rather than independently verified individual-tree damage severity. Further validation using independently assessed individual trees across all damage classes is required before the accuracy of operational tree-level damage classification can be fully established.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/rs18234122/s1, Figure S1. Directional spatial-validation design using five contiguous X-coordinate bands; Figure S2. Classification performance of NDRE-only (M1) and NDRE–NRCT (M12) models under directional spatial cross-validation; Figure S3. Reciprocal edge-to-edge spatial transfer performance of M1 and M12; Figure S4. Predictor-distribution overlap under directional spatial validation; Figure S5. Fold-level comparison of spatially cross-validated classification performance for M1 and M12; Figure S6a. Predicted damage-class probabilities as functions of NDRE and NRCT for alternative logistic regression models; Figure S6b. Class-specific probability-calibration curves for alternative logistic regression models; Figure S6c. Spatial distribution of changes in predicted hail-damage classifications following relaxation of the proportional-odds assumption; Table S1. Five-band directional spatial-validation results; Table S2. Reciprocal edge-to-edge spatial transfer results; Table S3. Training–validation predictor-distribution overlap; Table S4. Assessment of the proportional-odds assumption for the NDRE–NRCT ordinal model (M12); Table S5. Comparison of probabilistic prediction performance and calibration for selected ordinal regression models; Table S6. Paired classification-performance differences between standalone spectral indices and their combinations with LiDAR-derived LAI; Table S7a. Comparison of LAI estimators on the common-tree population; Table S7b. Effect of LAI estimator selection on spatially cross-validated classification performance; Table S7c. Classification performance of direct gap-fraction predictors; Table S7d. Sensitivity of gap fraction to canopy-height thresholds and LiDAR return-selection methodology; Table S8a. Global and predictor-specific tests of the proportional-odds assumption; Table S8b. Boundary-specific regression coefficients and tests of coefficient differences; Table S8b(i). Coefficient estimates from cumulative binary logistic regression models; Table S8b(ii). Differences between boundary-specific coefficients; Table S8c. Coefficients for proportional-odds, partial proportional-odds and multinomial regression model; Table S8d. Paired spatial-bootstrap comparison of out-of-fold model performance; Table S8e(i). Overall classification and probabilistic performance by spatial validation fold; Table S8e(ii). Class-specific discrimination statistics by spatial validation fold; Table S8f(i). Overall classification and probabilistic performance under spatial cross-validation; Table S8f(ii). Class-specific discrimination under spatial cross-validation; Table S8g. Class-specific probability-calibration results under spatial cross-validation; Table S9a. Sensitivity of NRCT to alternative normalization bounds; Table S9b. Sensitivity of NRCT to LiDAR-based canopy masking; Table S9b(i). Effect of canopy masking on crown-level NRCT and classification performance; Table S9b(ii). Paired differences in spatially cross-validated balanced accuracy under alternative canopy-masking configurations. The Supplementary Information provides additional methodological details, diagnostic analyses, and supporting results. It includes directional and fold-level spatial cross-validation results; assessments of predictor-distribution overlap and spatial transferability; comparisons of spectral indices and their integration with LiDAR-derived LAI and canopy gap fraction; sensitivity analyses of LiDAR return-selection methods, canopy-height thresholds, and thermal normalisation procedures; and detailed evaluations of the proportional-odds assumption and alternative ordinal and multinomial regression models. Supporting tables and figures report model coefficients, classification performance, spatial-bootstrap confidence intervals, class-specific discrimination, probability calibration, and the spatial distribution of changes in predicted damage severity.

Author Contributions: Conceptualisation, A.F., D.P.; methodology, A.F.; software, A.F., P.S.M.S.; validation, A.F., D.P.; formal analysis, A.F., P.S.M.S.; investigation, A.F., D.P.; data curation, A.F., P.S.M.S.; writing—original draft preparation, A.F.; writing—review and editing, D.P., P.S.M.S.; project administration, D.P.; funding acquisition, D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by the project Green Triangle Forest Health and Biosecurity Program managed by the Forestry Centre of Excellence, Adelaide University, Mount Gambier, Contract ID 114949, Project ID 60150472.

Data Availability Statement: Data are available from the corresponding author upon reasonable request.

Acknowledgments: We are grateful to Steven Andriolo of EyeSky for conducting the drone operations in Victoria, and to the Green Triangle Forest Health Group and its supporting companies, Australian Bluegum Plantations (ABP), HVP Plantations (HVP), Green Triangle Forest Products (GTFP), Midway Limited, PF Olsen (PFO), Timberlands Pacific and OneFortyOne (OFO) for supporting this study, and to Timberlands Pacific for access to the site. The authors also thank HxGN SmartNet for providing access to their GNSS NTRIP correction services, which were utilised during data collection, free of charge for educational research purposes. During the preparation of this manuscript, the authors used ChatGPT (version GPT 6) to convert MATLAB 2026a-generated and author-created graphics into publication-ready figures and to assist with language review. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

NDRE	Normalised difference red edge
NRCT	Normalised relative canopy temperature
NDVI	Normalised difference vegetation index
GLI	Green leaf index
CCCI	Canopy chlorophyll content index
ICCI	Integrated canopy condition index
$ICCI_G$	Integrated canopy condition index (green)
LAI	Leaf area index
CC	Canopy cover
GF	Gap fraction
UAV	Unmanned aerial vehicle
ROC	Receiver operating characteristic
AUC	Area under the ROC curve
RGB	Red, green, and blue
SD	Standard deviation
CV	Cross-validation
FWHM	Full width at half maximum
GNSS	Global navigation satellite system
NTRIP	Networked Transport of RTCM via Internet Protocol
RTK	Real-time kinematic
RTCM	Radio Technical Commission for Maritime Services

References

1. Allen, J.T.; Giammanco, I.M.; Kumjian, M.R.; Punge, H.J.; Zhang, Q.; Groenemeijer, P.; Kunz, M.; Ortega, K. Understanding hail in the Earth system. *Rev. Geophys.* **2020**, *58*, e2019RG000665. [CrossRef]
2. Cremer, K.W. Hail damage in Australian pine plantations I. Nature and extent of damage. *Aust. For.* **1984**, *47*, 103–114. [CrossRef]
3. Cremer, K.W. Nature and impact of damage by wind, hail and snow in Australia's pine plantations. *Aust. For.* **1984**, *47*, 28–38. [CrossRef]

4. Ghosh, S.J.L.; Innes, J.L.; Hoffmann, C. Observer Variation as a Source of Error in Assessments of Crown Condition Through Time. *For. Sci.* **1995**, *41*, 235–254. [CrossRef]
5. Ecke, S.; Dempewolf, J.; Frey, J.; Schwaller, A.; Endres, E.; Klemmt, H.-J.; Tiede, D.; Seifert, T. UAV-Based Forest Health Monitoring: A Systematic Review. *Remote Sens.* **2022**, *14*, 3205. [CrossRef]
6. Minko, G. Long term effects of hail damage on radiata pine. *Aust. For.* **1979**, *42*, 168–171. [CrossRef]
7. Wingfield, M.J.; Slippers, B.; Barnes, I.; Duong, T.A.; Wingfield, B.D. The Pine Pathogen Diplodia sapinea: Expanding Frontiers. *Curr. For. Rep.* **2025**, *11*, 2. [CrossRef]
8. Raupach, T.H.; Soderholm, J.S.; Warren, R.A.; Sherwood, S.C. Changes in hail hazard across Australia: 1979–2021. *npj Clim. Atmos. Sci.* **2023**, *6*, 143. [CrossRef]
9. Fraser, B.T.; Congalton, R.G. Monitoring Fine-Scale Forest Health Using Unmanned Aerial Systems (UAS) Multispectral Models. *Remote Sens.* **2021**, *13*, 4873. [CrossRef]
10. Ecke, S.; Stehr, F.; Frey, J.; Tiede, D.; Dempewolf, J.; Klemmt, H.-J.; Endres, E.; Seifert, T. Towards operational UAV-based forest health monitoring: Species identification and crown condition assessment by means of deep learning. *Comput. Electron. Agric.* **2024**, *219*, 108785. [CrossRef]
11. Lassalle, G. Monitoring natural and anthropogenic plant stressors by hyperspectral remote sensing: Recommendations and guidelines based on a meta-review. *Sci. Total Environ.* **2021**, *788*, 147758. [CrossRef] [PubMed]
12. Berger, K.; Machwitz, M.; Kycko, M.; Kefauver, S.C.; Van Wittenberghe, S.; Gerhards, M.; Verrelst, J.; Atzberger, C.; van der Tol, C.; Damm, A.; et al. Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. *Remote Sens. Environ.* **2022**, *280*, 113198. [CrossRef] [PubMed]
13. Bannari, A.; Morin, D.; Bonn, F.; Huete, A.R. A review of vegetation indices. *Remote Sens. Rev.* **1995**, *13*, 95–120. [CrossRef]
14. Zeng, Y.; Hao, D.; Huete, A.; Dechant, B.; Berry, J.; Chen, J.M.; Joiner, J.; Frankenberg, C.; Bond-Lamberty, B.; Ryu, Y.; et al. Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nat. Rev. Earth Environ.* **2022**, *3*, 477–493. [CrossRef]
15. Qiao, L.; Tang, W.; Gao, D.; Zhao, R.; An, L.; Li, M.; Sun, H.; Song, D. UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. *Comput. Electron. Agric.* **2022**, *196*, 106775. [CrossRef]
16. Jones, H.G. Application of thermal imaging and infrared sensing in plant physiology and eco-physiology. *Adv. Bot. Res.* **2004**, *41*, 107–163. [CrossRef]
17. Maes, W.H.; Steppe, K. Estimating evapotranspiration and drought stress with ground-based thermal remote sensing in agriculture: A review. *J. Exp. Bot.* **2012**, *63*, 4671–4712. [CrossRef] [PubMed]
18. Jackson, R.D.; Idso, S.B.; Reginato, R.J.; Pinter, P.J., Jr. Canopy temperature as a crop water stress indicator. *Water Resour. Res.* **1981**, *17*, 1133–1138. [CrossRef]
19. Smigaj, M.; Agarwal, A.; Bartholomeus, H.; Decuyper, M.; Elsherif, A.; de Jonge, A.; Kooistra, L. Thermal infrared remote sensing of stress responses in forest environments: A review of developments, challenges, and opportunities. *Curr. For. Rep.* **2024**, *10*, 56–76. [CrossRef]
20. Jones, H.G. Thermal Imaging and Infrared Sensing in Plant Ecophysiology. In *Advances in Plant Ecophysiology Techniques*; Springer: Heidelberg, Germany, 2018; pp. 135–151.
21. Mitchell, S.J. Wind as a natural disturbance agent in forests: A synthesis. *Forestry* **2013**, *86*, 147–157. [CrossRef]
22. Lefsky, M.A.; Cohen, W.; Acker, S.; Parker, G.; Spies, T.; Harding, D. Lidar Remote Sensing of the Canopy Structure and Biophysical Properties of Douglas-Fir Western Hemlock Forests. *Remote Sens. Environ.* **1999**, *70*, 339–361. [CrossRef]
23. Duncanson, L.I.; Niemann, K.O.; Wulder, M.A. Estimating forest canopy height and terrain relief from GLAS waveform metrics. *Remote Sens. Environ.* **2010**, *114*, 138–154. [CrossRef]
24. Zhou, M.; Li, C.; Li, Z. Extraction of individual tree attributes using ultra-high-density point clouds acquired by low-cost UAV-LiDAR in Eucalyptus plantations. *Ann. For. Sci.* **2025**, *82*, 20. [CrossRef]
25. Zhang, Y.; Tan, Y.; Xu, X.; You, H.; Onda, Y.; Gomi, T. Individual tree branch and leaf metrics extraction in dense plantation scenario through the fusion of drone and terrestrial LiDAR. *Comput. Electron. Agric.* **2025**, *232*, 110070. [CrossRef]
26. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **1994**, *38*, 113–125. [CrossRef]
27. Finn, A.; Younger, J.; Skelton, P.S.M.; Peters, S.; O'Hehir, J.; Turner, D.; Lucieer, A. Extracting Value from Fused Aerial and Terrestrial LiDAR Scans. *Remote Sens.* **2026**, *18*, 2644. [CrossRef]
28. Fang, H.; Baret, F.; Plummer, S.; Schaepman-Strub, G. An overview of global leaf area index (LAI): Methods, products, validation, and applications. *Rev. Geophys.* **2019**, *57*, 739–799. [CrossRef]
29. Tian, L.; Qu, Y.; Qi, J. Estimation of forest LAI using discrete airborne LiDAR: A review. *Remote Sens.* **2021**, *13*, 2408. [CrossRef]
30. Solberg, S. Mapping gap fraction, LAI and defoliation using various ALS penetration variables. *Int. J. Remote Sens.* **2010**, *31*, 1227–1244. [CrossRef]
31. Bussotti, F.; Potočić, N.; Timmermann, V.; Lehmann, M.M.; Pollastrini, M. Tree crown defoliation in forest monitoring: Concepts, findings, and new perspectives for a physiological approach in the face of climate change. *Forestry* **2024**, *97*, 194–212. [CrossRef]

32. Iheaturu, C.J.; Hepner, S.; Batchelor, J.L.; Agonvonon, G.A.; Akinyemi, F.O.; Wingate, V.R.; Speranza, C.I. Integrating UAV LiDAR and multispectral data to assess forest status and map disturbance severity in a West African forest patch. *Ecol. Inform.* **2024**, *84*, 102876. [CrossRef]
33. Savinelli, B.; Tagliabue, G.; Vignali, L.; Garzonio, R.; Gentili, R.; Panigada, C.; Rossini, M. Integrating Drone-Based LiDAR and Multispectral Data for Tree Monitoring. *Drones* **2024**, *8*, 744. [CrossRef]
34. El-Shikha, D.M.; Barnes, E.M.; Clarke, T.R.; Hunsaker, D.J.; Haberland, J.A.; Pinter, P.J., Jr.; Waller, P.M.; Thompson, T.L. Remote Sensing of Cotton Nitrogen Status Using the Canopy Chlorophyll Content Index (CCCI). *Trans. ASABE* **2008**, *51*, 73–82. [CrossRef]
35. Fitzgerald, G.; Rodriguez, D.; O'Leary, G. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). *Field Crops Res.* **2010**, *116*, 318–324. [CrossRef]
36. Iglhaut, J.; Cabo, C.; Puliti, S.; Piermattei, L.; O'Connor, J.; Rosette, J. Structure from motion photogrammetry in forestry: A review. *Curr. For. Rep.* **2019**, *5*, 155–168. [CrossRef]
37. Pingel, T.J.; Clarke, K.C.; McBride, W.A. An improved simple morphological filter for the terrain classification of airborne LIDAR data. *ISPRS J. Photogramm. Remote Sens.* **2013**, *77*, 21–30. [CrossRef]
38. Gonzalez-Benecke, C.A.; Fernández, M.P.; Gayoso, J.; Pincheira, M.; Wightman, M.G. Using Tree Height, Crown Area and Stand-Level Parameters to Estimate Tree Diameter, Volume, and Biomass of Pinus radiata, Eucalyptus globulus and Eucalyptus nitens. *Forests* **2022**, *13*, 2043. [CrossRef]
39. Yang, J.; Kang, Z.; Cheng, S.; Yang, Z.; Akwensi, P.H. An Individual Tree Segmentation Method Based on Watershed Algorithm and Three-Dimensional Spatial Distribution Analysis From Airborne LiDAR Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1055–1067. [CrossRef]
40. Phillips, C.L. Forest health surveillance in South Australia. *Aust. For.* **2008**, *71*, 196–201. [CrossRef]
41. McConnell, T.J.; Johnson, E.W.; Burns, B. *A Guide to Conducting Aerial Sketchmapping Surveys*; U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team: Fort Collins, CO, USA, 2000.
42. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [CrossRef]
43. Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. Block CV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **2019**, *10*, 225–232. [CrossRef]
44. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
45. Ploton, P.; Mortier, F.; Réjou-Méchain, M.; Barbier, N.; Picard, N.; Rossi, V.; Dormann, C.; Cornu, G.; Viennois, G.; Bayol, N.; et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **2020**, *11*, 4540. [CrossRef] [PubMed]
46. Brant, R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics* **1990**, *46*, 1171–1178. [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.