Supplementary Material S1

Model implementation details

This section provides the full implementation details of the Sentinel-2 U-Net classifier underlying the flagging framework. Stand-level aggregation and agreement assessment, the flag definition and class-precision stratification, and the validation design, including the control group and blinded image interpretation, are described in the main text (Sections 2.2.3–2.4), and the per-class precision results underlying the high-precision/low-precision stratification are reported in Section 3.1. Details not already given in the main text are reported here. The workflow was adapted from Lee et al. (2025) [1]. The predictor set and class scheme were modified for the present study.

S1.1 Satellite imagery and compositing

The input imagery was the Surface Reflectance Harmonized product (COPERNICUS/S2_SR_HARMONIZED) of Sentinel-2 (ESA). Ten bands spanning the visible, red-edge, near-infrared, and shortwave-infrared regions were selected for their sensitivity to vegetation spectral response: B2, B3, B4, B5, B6, B7, B8, B8A, B11, and B12. Three seasonal composites were built from the April, September, and November 2025 imagery to capture leaf-out, peak canopy, and pre-abscission states, respectively.

Compositing was performed on Google Earth Engine [2]. For each seasonal composite, only scenes from the corresponding month with cloud cover below 50% were retained; pixels flagged as cloud, cloud shadow, cirrus, or snow were masked using the Scene Classification Layer (SCL) [3]. The pixel-wise median was then computed to suppress residual cloud and transient variability. Pixels with no valid observations were excluded from the analysis. Bands at 20 m native resolution were resampled to 10 m by bilinear interpolation, and the seasonal composites were clipped to the 4 × 4 km fishnet cells.

S1.2 Vegetation indices

For each seasonal composite, seven vegetation indices were derived to quantify chlorophyll content, canopy vigor, and soil-background effects by combining the visible, red-edge, and near-infrared regions. The formulas below use standard Sentinel-2 band assignments.

Table S1. Vegetation indices computed per seasonal composite.

Index	Formula (Sentinel-2 bands)	Reference
NDVI	(B8 − B4) / (B8 + B4)	Rouse et al. (1974) [4]
GNDVI	(B8 − B3) / (B8 + B3)	Gitelson et al. (1996) [5]
RVI	B8 / B4	Jordan (1969) [6]
NDRE	(B8 − B5) / (B8 + B5)	Gitelson & Merzlyak (1994) [7]
CIre	(B7 / B5) − 1	Gitelson et al. (2003) [8]
MCARI	[(B5 − B4) − 0.2(B5 − B3)] × (B5 / B4)	Daughtry et al. (2000) [9]
SAVI	1.5 × (B8 − B4) / (B8 + B4 + 0.5)	Huete (1988) [10]

S1.3 GLCM texture features

Seven Gray-Level Co-occurrence Matrix (GLCM) texture features were computed on the near-infrared band (B8) using a 7×7 sliding window [11]. Within each window, GLCMs were constructed for four directions (0°, 45°, 90°, 135°) at a distance of 1 pixel and averaged across directions to enforce directional isotropy.

Table S2. GLCM texture features (B8, 7×7 window, distance 1, four directions averaged).

Feature	Description
Contrast	Local intensity variation between neighboring pixels
Dissimilarity	Linear measure of local gray-level difference
Homogeneity	Closeness of the GLCM distribution to its diagonal
Energy	Textural uniformity (square root of ASM)
Correlation	Linear gray-level dependency between pixel pairs
ASM	Angular Second Moment; textural orderliness
Entropy	Randomness / disorder of the gray-level distribution

S1.4 Topographic channels

Four time-invariant topographic channels were appended: elevation (DEM), slope, and the sine and cosine components of aspect. Slope and aspect were derived from the National Geographic Information Institute DEM in 2023 [12]. To avoid the discontinuity of aspect at 0°/360° hindering learning near boundaries, sin(aspect) and cos(aspect) were computed at native resolution—yielding two smooth channels bounded in [−1, 1]—before reprojection to the 10 m analysis grid. Flat areas (aspect = −1) and NoData were set to 0 in both channels. DEM and slope were resampled by bilinear interpolation, and the pre-decomposed sine/cosine channels were aligned to the per-cell label raster likewise by bilinear interpolation.

S1.5 Label preparation and split allocation

Labels were built on the 2025 1:5,000 FGIS digital forest type map [13]. Protected areas and infrastructure-related features were masked, coniferous–broadleaf mixed stands were removed, and stands smaller than 50 pixels (0.5 ha) were discarded, as described in the main text (Sections 2.1 and 2.2.1); mixed-forest polygons were assigned NoData. Before split allocation, 4 × 4 km fishnet cells in which 10% or more of the area had changed since map production owing to harvesting, wildfire, or other disasters, or was affected by residual haze or cloud, were removed from the candidate cells based on visual inspection of the imagery. Stands in these cells were excluded from model training and test-set agreement assessment but remained evaluable and eligible for flagging (main text, Section 2.2.1).

Cells were assigned to the training, validation, and test splits by the rarity-ordered procedure described in the main text (Section 2.2.1). Each cell was assigned to one split, and the selected cell list was stored together with the 4 × 4 km fishnet polygons. The same 14 test cells, containing 5,797 test stands, were used in all three main replicates; model initialization and training-cell selection varied. Table S3 reports the allocation used for the first replicate and subsequent analyses. The predictor ablation held the full cell allocation fixed while varying initialization seeds. Target shares refer to per-class pixel counts; forest areas are reported separately in Table S3.

Table S3. Allocation of fishnet cells to splits for the first replicate and subsequent analyses. Target shares are per-class pixel targets (main text, Section 2.2.1).

Split	Target pixel share	Cells	Forest area (ha)
Training	24%	55	58,165
Validation	3%	10	8,465
Test	3%	14	14,069
Total	30%	79	80,699

S1.6 Input-tensor construction and normalization

For each season (April, September, November), 10 spectral bands, 7 vegetation indices, and 7 GLCM textures were combined into 24 channels; the three seasons were concatenated to 72 channels, to which the 4 topographic channels were appended, yielding a (76, 400, 400) tensor per 4 × 4 km cell.

Region-wise percentile normalization mitigated channel-wise distributional differences: per-channel lower/upper percentiles were computed from random samples within a region, and each channel was linearly mapped to [0, 1] using a 2nd–98th percentile transform. An IQR-based outlier filter (k = 3) reduced the influence of extreme values such as those in GLCM. Normalization statistics were stored per region in JSON so the identical transform could be reproduced at inference.

S1.7 Network architecture

A standard U-Net semantic-segmentation architecture was used. The encoder comprises four downsampling stages, each with two convolution blocks (3×3 convolution, batch normalization, ReLU) followed by 2×2 max pooling; feature channels grow as 64 → 128 → 256 → 512, expand to 1,024 at the bridge, and reduce back to 512. The symmetric decoder restores resolution with 2×2 transposed convolution (stride 2), concatenates encoder features of equal resolution via skip connections, and applies two convolution blocks per stage. Class logits are produced by a 1×1 convolution.

The model operates in a fixed, region-independent label space of 16 output channels (class 0 = background/NoData; classes 1–15 = species/other); only the classes present in a region (active classes) are activated for training and evaluation. For the Seoul–Gyeonggi region, the active classes correspond to the seven consolidated classes defined in the main text (Section 2.1, Table 1).

S1.8 Training configuration

Table S4. Training hyperparameters and configuration.

Setting	Value
Loss	Class-weighted cross-entropy; background and NoData pixels (label 0) excluded from the loss by mapping them to ignore_index = −100
Class weights	$w(c) = N_{total} / (K \times n_c)$, K = number of active classes, n_c = pixels in class c
Optimizer	Adam [14], initial learning rate 0.001
LR schedule	ReduceLROnPlateau (factor 0.5, patience 5, floor 1×10^{-6}) on validation accuracy
Gradient clipping	Global norm clipped to 1.0
Numerical precision / batching	Mixed precision (AMP); effective batch size 8 (batch 2 × accumulation 4)
Early stopping	Stop if validation accuracy does not improve for 10 epochs; max 100 epochs
Model selection	Best validation-accuracy weights saved as final model
Replicates	Three runs with different initialization seeds and training-cell selections; the test cells were fixed. Main-text metrics are mean ± SD (n = 3).
Framework	PyTorch [15]

Training patches of 256×256 pixels were extracted from each 4 × 4 km cell (400×400 pixels) at the four corners; patches with a valid-pixel ratio below 0.3 were discarded. Five deterministic augmentations (original, vertical flip, horizontal flip, +90° and −90° rotation) were applied to the training data. To reduce label noise, training labels underwent per-class 1-pixel boundary erosion and removal of connected components smaller than 5 pixels, with the affected pixels ignored in the loss; because erosion is per class, boundary pixels between two adjacent species are excluded on both sides, preventing forced assignment of mixed boundary pixels. The imagery itself was not masked and original reflectance was preserved, because the non-forest surrounding context aids stand-boundary discrimination and avoids a train/prediction distribution shift.

S1.9 Prediction

Prediction used the sliding-window soft vote described in the main text (Section 2.2.2): tiles of 256×256 pixels were extracted at a stride of 128 pixels, and class probabilities from overlapping tiles were averaged before the arg max. Three rasters were written per cell: the predicted class; the pixel confidence, defined as the averaged softmax probability of the predicted class; and the pixel decision margin, defined as the difference between the highest and second-highest averaged probabilities. The aggregation of these pixel-level outputs to the stand-level dominant class, its areal share (dom_share), and the mean pixel confidence (conf) is given in Section 2.2.3 of the main text; the flag definition and thresholds, the class-precision stratification, and the validation design, including the control group and blinded image interpretation, are given in Sections 2.3 and 2.4.

References

1. Lee, H.; Lee, C.; Woo, H.; Choi, S.-E. Optimal training sample sizes for U-Net-based tree species classification with Sentinel-2 imagery. *Forests* **2025**, *16*, 1718. https://doi.org/10.3390/f16111718.
2. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.
3. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In *Proceedings of the Image and Signal Processing for Remote Sensing XXIII*, Warsaw, Poland, 11–14 September **2017**; Volume 10427, 1042704. https://doi.org/10.1117/12.2278218.
4. Rouse, J.W., Jr.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium*, Greenbelt, MD, USA, 10–14 December **1973**; NASA SP-351; NASA: Washington, DC, USA, **1974**; Volume I, pp. 309–317.
5. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7.
6. Jordan, C.F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology* **1969**, *50*, 663–666. https://doi.org/10.2307/1936256.
7. Gitelson, A.; Merzlyak, M.N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *J. Photochem. Photobiol. B* **1994**, *22*, 247–252. https://doi.org/10.1016/1011-1344(93)06963-4.
8. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. https://doi.org/10.1078/0176-1617-00887.
9. Daughtry, C.S.T.; Walthall, C.L.; Kim, M.S.; Brown de Colstoun, E.; McMurtrey, J.E. Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sens. Environ.* **2000**, *74*, 229–239. https://doi.org/10.1016/S0034-4257(00)00113-9.
10. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.
11. Haralick, R.M.; Shanmugam, K.; Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *SMC-3*, 610–621. https://doi.org/10.1109/TSMC.1973.4309314.
12. National Geographic Information Institute. National Land Information Platform. Available online: https://map.ngii.go.kr/ (accessed on 21 May **2025**).
13. Korea Forest Service. Forest Geographic Information Service. Available online: https://map.forest.go.kr/forest/ (accessed on 27 February **2026**).
14. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, 7–9 May **2015**.
15. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An imperative style, high-performance deep learning library. In Advances in *Neural Information Processing Systems* 32; **2019**; pp. 8024–8035. Available online: https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html (accessed on 26 October 2026).