Geospatial tree species prediction using multi-view drone imagery and computer vision

Amritha S. Pallavoor[1*], David J. Russell[1], Derek J. N. Young[1]

[1]Department of Plant Sciences, University of California Davis

*Corresponding author: aspallavoor@ucdavis.edu

Keywords– Species prediction, geospatial, computer vision, multi-view, tree, drone, UAV

Abstract

Imagery from uncrewed aerial vehicles (“drones”) is an increasingly popular modality for understanding forests at a large scale due to its relatively low cost and ability to be collected on demand. Computational tools to identify the location of individual trees and their species can inform forest management efforts, such as prioritizing regions to thin to reduce wildfire risk. Previous work on species prediction has relied on imagery with a single top-down view produced from an *orthomosaic*, which contains image artifacts and limited information about the side views of the tree. In this work, we used the high resolution raw drone imagery to train machine learning models, and leverage its highly overlapping nature to capture multiple views of each tree for species prediction. We precisely geolocated individual drone images and estimated 3D geospatial models of each stand using *photogrammetry*. These photogrammetry products were used to derive Canopy Height Models (CHM) to detect tree crowns and identify individual treetops through a geometry-based local maxima algorithm. Then, we spatially aligned the drone detected trees to the field data, matched trees between the modalities, and assigned field-observed species to the detected trees. Using *Geograypher,* we cropped each individual tree crown from every raw image it appears in and used this dataset to train an EfficientNetV2 species prediction model. For unseen data, we detected trees, generated species predictions independently across image views, and assigned each tree the most frequently predicted species class. In this workflow, we used data collected from 107 field plots comprising 9,238 matched trees. The multi-view drone image prediction workflow achieved a macro-averaged recall of 79% compared to the baseline of 63% for the orthomosaic approach. Our results demonstrate that leveraging the inherent multi-view nature of raw drone images and aggregating predictions at the individual tree level improves species classification accuracy across a geospatially diverse set of conifer forests in the western U.S. This workflow enables scalable tree-level species mapping and can support downstream applications in forest inventory and management.

Introduction

Understanding the tree species composition of forests is essential for many ecological and natural resource management applications, including assessing and predicting the impacts of disturbances such as drought and wildfire, monitoring and projecting carbon stocks, and planning forest restoration approaches (Schadauer et al., 2024). Many of these applications require knowledge of species composition across large areas. However, the manual field-based surveys that traditionally produce such data are labor-intensive and time-consuming. To overcome this constraint, forest ecologists and managers sometimes turn to aerial and satellite imagery. Widely available and relatively low-cost uncrewed aerial vehicles (UAVs), or "drones", can capture imagery with a spatial resolution of ~1-5 cm and have been demonstrated to enable accurate fine-scale predictions of vegetation cover across large landscapes (Russell et al., 2025). Models built on multispectral satellite imagery can yield predictions of species abundance across space (Ohmann & Gregory, 2002; *TreeMap 2016*, n.d.), but their accuracy is constrained by the fact that satellite pixels (generally 30 m for existing products) (a) often do not correspond to individual tree crowns but rather combinations of one or more tree crowns, other non-tree vegetation, and/or non-vegetative ground cover, and (b) are too coarse to capture morphological species cues in individual tree crowns (Hsieh & Lee, 2000; Onishi & Ise, 2021). Even the highest-resolution datasets available from public (~10 m) and commercial (~1 m) satellite imagery sources suffer from these limitations to varying degrees. Aerial imaging platforms offer the potential to collect much higher resolution imagery–often hundreds or thousands of pixels per tree crown–enabling approaches that distinguish species based on their morphological characteristics in much the same way field biologists classify species by sight.

Early work on automated species prediction relied on traditional machine learning approaches such as support vector machine (SVM) and random forest (RF) models (Immitzer et al., 2012; Raczko & Zagajewski, 2017; Shang & Chisholm, 2014). However, these approaches are designed for tabular data, and applying them to images requires treating each pixel as an independent data point and applying *a priori* preprocessing steps ("feature extraction") to compute information describing the spatial/image context of each pixel or object (e.g., standard deviation of brightness in the neighboring 1 m, or mean intensity per spectral band of all pixels within a tree's crown). Recent advancements in deep learning and computer vision are more "image-native," have reduced the need for manual feature extraction, and can instead learn the most relevant ways to represent object appearance and neighborhood context from the training data (Krizhevsky et al., 2012; LeCun et al., 2015). Convolutional Neural Networks (CNNs) are capable of learning distinctive features of different image categories through hierarchical representations of visual elements progressing from low-level patterns such as edges and textures, to high-level features specific to each target class (Zeiler & Fergus, 2013; Schiefer et al., 2020). This enables CNNs to capture the subtle visual differences between tree species from drone imagery in ways that more conventional ML methods cannot.

Recent approaches have investigated computer vision models for tree species classification from drone imagery. EfficientNet, Vision Transformer (ViT), and You Only Look Once (YOLO) models have been trained to achieve F1 scores greater than 0.9 across eight species (Huang et

al., 2024), while other CNNs similarly demonstrate strong performance (Onishi & Ise, 2021). Yet, many studies have largely been confined to single sites, limiting their generalizability and prompting the need for models trained across diverse sites and time periods. Weinstein et al., (2023) began addressing this gap by developing a multi-temporal hierarchical CNN model to predict 14 tree species, including rare species with <1% frequency, at a single NEON site in Florida, achieving over 75% accuracy. Building on this, the approach was scaled to 81 species across 24 NEON sites in the U.S., combining multi-year airborne RGB (10cm), hyperspectral (1m), and LiDAR data to generate predictions for 100 million individual trees with 79% average accuracy (Weinstein et al., 2024).

Standard drone surveys are typically conducted by flying a UAV along a dense series of parallel transects capturing images that are highly overlapping with one another from front to back (along transects) and side to side (across transects). In this way, the drone captures dozens of distinct views of every object on the ground from multiple viewpoints. Most of the existing tree species prediction approaches rely on a single top-down orthomosaic view that is generated from this multi-view imagery. However, this top-down view provides limited information with which to infer a tree's species and ignores the wealth of information available from the multiple oblique views of each tree in the raw drone images. Furthermore, the orthorectification process inherent in orthomosaic production can degrade image quality by introducing warping artifacts and distortion (Russell et al., 2024). Additionally, because deep learning models generally require large datasets for effective training, traditional methods rely on synthetic data augmentation techniques like rotation, cropping, and brightness variations. Multi-view datasets solve this data requirement naturally by providing genuinely different perspectives of the same tree which leads to improved species classification accuracy (Liu & Abd-Elrahman, 2018; Russell et al., 2024).

Although they can achieve high image-level accuracy (e.g., Santos et al., 2019), many tree detection and species classification approaches do not map predictions back to geographic coordinates. However, understanding the geospatial locations of individual trees by species is essential for many forest ecology and management applications. Furthermore, understanding the geospatial locations of trees in raw images is necessary to associate multiple views of the same tree, which is in turn necessary to split the images into train and test sets at the tree level. Ignoring these associations and instead splitting datasets randomly by image risks the same tree appearing in both the training and evaluation sets, artificially inflating accuracy estimates. Finally, having species predictions from multiple views of the same tree allows for aggregating predictions to obtain a single tree-level prediction with potentially much higher confidence than any individual view can provide.

Until very recently, no work had leveraged multiple views of a tree aggregated to a geospatial location for a species prediction task. A solution to these technical tasks was developed by Russell et al., (2024), who demonstrated it on 773 trees across 4 sites using a computer vision segmentation model. However, this contribution did not provide a framework for deploying these solutions at scale or evaluate different data subsetting and class aggregation techniques for training and inferencing from a computer vision model. Furthermore, it employed a

segmentation model, as opposed to a classification model, which presents challenges for interpretation and scaling. Here, we demonstrate an approach for integrating deep learning computer vision models with multi-view raw drone imagery datasets to predict tree species at the individual-tree level. We evaluate the approach using a dataset consisting of traditional manual ground-based inventory data and drone imagery from 107 plots (9,238 trees representing 8 of the most common species) spanning diverse environmental conditions across California yellow pine and mixed-conifer forests – a substantial increase in scale compared to previous multi-view tree species classification studies, with the potential for large accuracy gains. At the core of the workflow is a computer vision model coupled with a system for geometric reasoning that links individual trees in raw drone images to their geospatial locations.

Materials and Methods

Overview

We developed an automated workflow that integrates a photogrammetry-derived 3D mesh model (Russell et al., 2024), tree crown detection, and precise georeferencing to assign species labels derived from manual ground-based inventories to individual trees captured in the raw drone imagery. We used this pipeline to build a training and testing dataset for geospatial tree species prediction from raw drone imagery. This workflow processes both high-altitude nadir (top-down views) and low-altitude oblique imagery to leverage diverse views of each tree. We (a) quantified how models may benefit from aggregating predictions across different views through majority voting and (b) compared tree species prediction performance achievable between single-view orthomosaic imagery and several different collections of multi-view perspectives. We also (c) evaluated how different levels of taxonomic aggregation (before model training or after inference) affect classification accuracy and (d) explored whether regionally-specific models achieve higher accuracy.

Study sites and field reference data

We used drone data and co-located ground reference survey data from the Open Forest Observatory database (Observatory, 2026) from 107 forested sites in California, USA. The OFO database consists of datasets compiled from publicly available sources and contributed by individual researchers and groups.

The 107 focal sites in this study span a gradient of conifer-dominated forest types, from yellow pine and mixed-conifer forests (Safford & Stevens, 2017) to white fir (*Abies concolor*), red fir (*A. magnifica*), and lodgepole pine (*Pinus contorta*)-dominated stands, and they were selected to span representative gradients of stand structure attributes including mean tree size and tree density. Some sites had experienced moderate-severity wildfire within 10 years prior to drone imagery collection. Across all sites, the manual ground-based inventory was performed within nine years of the imagery collection and there was no significant disturbance between the two data collection dates at any plot.

The ground reference survey data we used contain the geospatial point locations of individual qualifying trees (specifically, their main stem at 1.3 m above ground level) along with attributes including species and size (height and/or stem diameter at 1.3 m above ground level) as assessed by technicians on the ground. Qualifying trees included all those that (depending on the dataset) met a minimum diameter at breast height (DBH) threshold of 10–25 cm and/or a minimum height threshold of 2–20 m. Across the plots we used, the spatial locations of each qualifying tree stem within specified plot bounds were determined largely by using a real-time kinematic (RTK) global navigation satellite system (GNSS) receiver (in sparser canopy conditions where a RTK fix enabled decimeter-level precision) and/or via “offset mapping” – determining the x-y offset of a given tree from a known geospatial survey point by measuring its distance and compass azimuth relative to that point. Tree stem diameters at breast height (DBHs) were measured by wrapping measuring tapes around tree stems at 1.3 m above ground level, and tree heights (if included in the dataset) were either measured using laser hypsometers or estimated based on measured diameter using local allometric equations. For datasets that did not include tree height data, we estimated height allometrically from measured DBH based on the following function:

$$height = 1.3 + \exp(-0.31365 + 0.846236 * \log(dbh)),$$

Where $height$ is tree height in meters and dbh is tree DBH in centimeters. We derived this formula via least squares linear regression of the parameters β_x in the following formula, using all included trees across all plots that had both DBH and height measurements:

$$\log(height-1.3) = \beta_0+\beta_1 \cdot \ln(dbh)$$

Drone imagery

We used 3-channel red-green-blue (RGB) broadband drone imagery from datasets collected by (depending on the dataset) a DJI Phantom 4 series drone (5472 × 3648 pixel 1" CMOS RGB sensor with an 84° diagonal field of view and 3.29 cm/pixel ground sampling distance at 120 m altitude) or a DJI Mavic 3 Multispectral drone (5280 × 3956 pixel 4/3" CMOS RGB sensor with an 84° field of view and 3.23 cm/pixel ground sampling distance at 120 m), both of which have mechanical shutters. Missions were flown with terrain tracking to maintain a constant altitude above ground level, but with differing altitudes, camera pitches, and image densities. We selected drone datasets (“missions”) in two different categories:

1. High Nadir (HN): Altitude above ground level (AGL) of 100-160m; camera pitch of 0-10 degrees up from directly downward (nadir); front overlap >= 90% and side overlap >= 80% or front and side overlap >= 85%.
2. Low Oblique (LO): Altitude above ground level (AGL) of 60-120m; camera pitch of 18-38 degrees up from directly downward (nadir); front overlap >= 70% and side overlap >= 60%.

The HN missions were intended to optimize geometric reconstruction and tree detection (Young et al. 2022), while the LO missions were intended to provide higher-resolution, oblique views of the trees, which we hypothesized would improve species prediction. To create imagery datasets for drone-based tree species prediction, we combined overlapping HN and LO missions. In cases where the geographic footprints of the HN and LO missions were not identical, we retained images in their intersection area.

For filtering missions and classifying them as HN or LO, , we used the nominal (programmed) mission image overlap values and mean drone-reported camera pitch values. Because actual mean flight altitude above ground level often deviated from the nominal flight altitude, we used the actual mean altitude above ground level inferred via photogrammetric processing of the resulting imagery (see below). To retain only those missions that reliably tracked terrain elevation, we required either (a) the Pearson correlation coefficient between the elevation of each image and the elevation of the ground beneath it to be > 0.75 or (b) the standard deviation of altitude to be < 12 m (to avoid excluding low-relief areas where small altitude deviations could produce low Pearson correlations).

We required both the HN and LO missions of a pair to be collected in the same calendar year.

Data pre-processing steps

We selected ground plots that were fully covered (including a 40 m buffer area around each ground plot) by both the HN and LO drone missions of a given mission pair.

Ground data and drone imagery collection dates of each pairing were required to be within nine years of each other. In some cases where drone mission footprints encompassed multiple ground plots, the same drone missions (though different parts of them) were paired with multiple ground plots. Similarly, in cases where a given ground plot was covered by multiple qualifying HN and/or LO drone missions, the HN mission with the smallest year difference relative to the ground survey was selected, with ties broken by the smallest date gap between the HN and LO missions. After the ground_plot–HN–LO sets were created, the geotagged raw drone images from the drone missions were subsetted to those collected within (above) the buffered ground plot. Our final analysis dataset included 107 ground plots but 133 ground_plot–HN–LO sets because some ground plots matched with multiple LO missions.

We processed raw images into a digital surface model, digital terrain model, mesh model, orthomosaic, and precisely estimated camera poses per image using a photogrammetry workflow based on Young et al. 2022. The workflow used Metashape version 2.2.0 (Agisoft LLC), run end-to-end via Automate Metashape version 0.3.0 (Young et al., 2024) using the configuration file template included in the data repository accompanying this paper.

Using the photogrammetry products, we produced a canopy height model (CHM) raster for each photogrammetry product by subtracting the DTM from the DSM. We used the resulting CHMs to generate drone-derived tree detections. First, we detected treetops using a local maximum filter

algorithm (Popescu & Wynne, 2004), parameterized to use a circular local maximum search window with a horizontal radius equal to 0.0325 times the height of the focal pixel plus 0.25 m. Next, the region around the treetop comprising the tree crown was delineated using a marker-controlled watershed segmentation approach (Meyer & Beucher, 1990) in which the treetop points detected in the previous step acted as seeds, with a constraint that only pixels above a minimum height threshold could be included in crown delineation. Treetop detection and tree crown delineation were performed using the `GeometricTreeTopDetector` and `GeometricTreeCrownDetector` classes in the Python package tree-detection-framework (Pallavoor et al., 2025).

Field and drone data are often not perfectly spatially co-registered due to differential error in the global navigation satellite system (GNSS) data used to record the locations of the field plot and/or drone images. Correctly aligning the field-reference data with the drone-derived tree detections is crucial to accurately assign field-derived tree species labels to drone-derived imagery and vice-versa. To align our datasets, we manually shifted all trees in a given field-based plot using a single, rigid plot-level x-y shift so that the “constellation” of field-mapped trees aligned best with the drone-derived canopy height model (sensu Young et al. 2022). If a shift producing clear alignment could not be found, the plot (along with its paired drone imagery) was excluded from subsequent analysis. (The total of 133 drone-field plot pairings reflects the count following exclusion of poorly aligned datasets.) For the remaining pairs, we next matched each shifted field tree to the corresponding drone-detected tree and vice-versa using the algorithm of Young et al. (2022). The algorithm greedily assigns field trees to drone trees based on proximity while satisfying a height similarity constraint, whereby the height of the drone tree must be within 0.5 and 1.5 times the height of the field tree. The detected tree crown polygons were then assigned attributes from the matched field tree, including species, live/dead status, and field-measured height.

To train a machine learning model on individual views of a tree, we must first determine each tree’s location in each image that views it. Following the approach of Russell et al. (2024), we used an open-source tool called Geographer (Russell, Sidhu, et al., 2025) to map between geospatial and image coordinates. This approach relies on a mesh model derived from photogrammetry to represent the scene, in combination with (a) the precise estimated position of the camera for each drone photo and (b) camera distortion parameters, both also derived from photogrammetry. Using this data as input, Geograypher uses 3D computer graphics techniques to determine which pixels in each image correspond to each geospatially segmented tree. Notably, this process is occlusion-aware, meaning that portions of trees that are blocked from view (e.g., by intervening trees) will be properly excluded. This approach allows us to obtain dozens of labeled views of each tree, especially critical side views, and avoids relying on imagery degraded by the orthomosaicing process.

Using the rendered tree instance IDs in each image, we extracted cropped views of each tree corresponding to the area defined by its mask while excluding background (non-focal tree) pixels by assigning them a 50% gray color (Fig. 1b-1e). We discarded all crops with size below 250x250 pixels to exclude views that were too incomplete or too distant to contain meaningful

detail. This size filter disproportionately removed HN views and led to complete removal of all nadir views for many trees (only 1311 trees out of 1896 in the test set had at least one retained nadir view), likely because HN images capture smaller crown areas (in image pixel dimensions) due to their high-altitude (distant) and top-down (minimal profile) perspective.

To compare the multi-view model against a single-view orthomosaic-based approach, we also developed a computer vision model training and testing dataset using tree crown images cropped from the orthomosaic of every plot. As with the raw drone images, we set the background pixels in the orthomosaic crops to 50% gray (Fig. 1a).

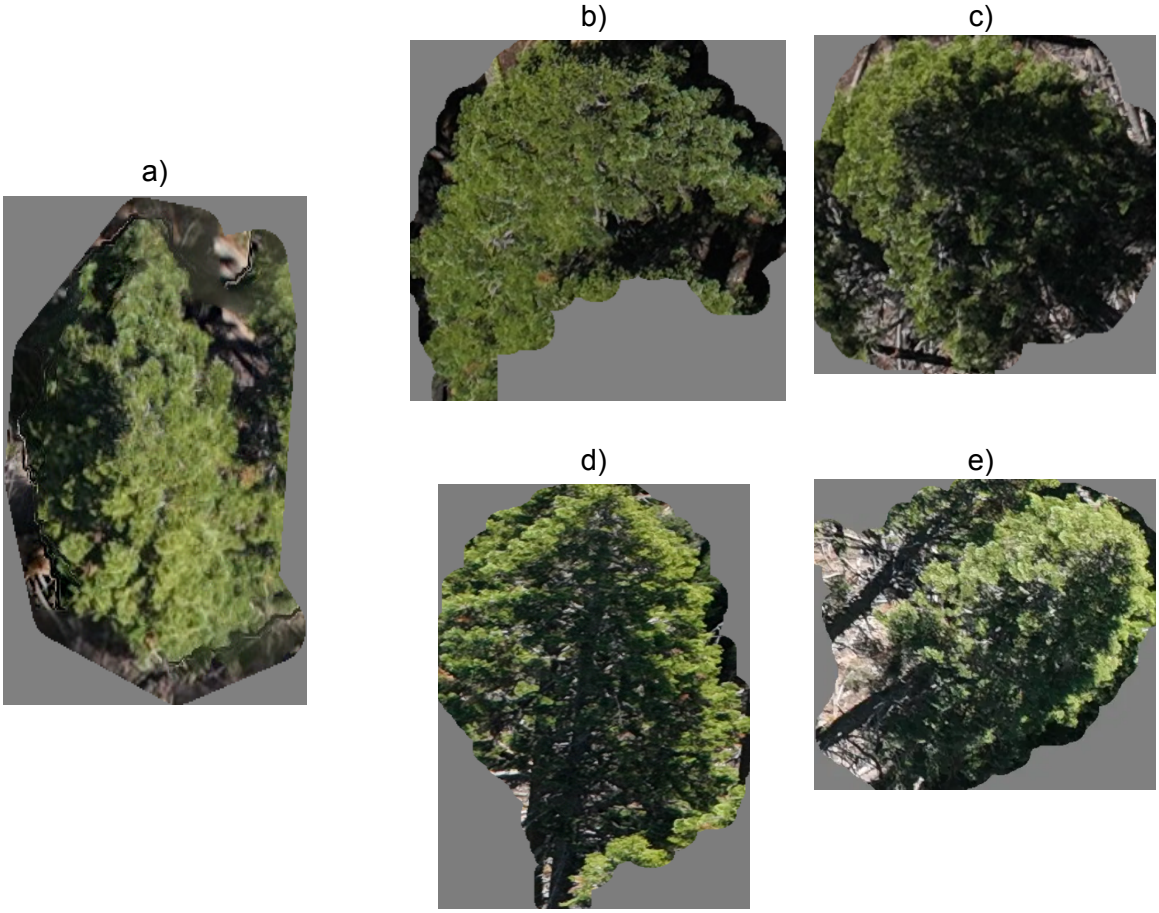

Figure 1: Preprocessed tree crops for a tree of class *Abies magnifica* (ABMA). (a) Crop from the orthomosaic. (b–c) High-nadir (HN) images. (d–e) Low-oblique (LO) images.

Dead or live tree classification

To filter dead trees from the species classification dataset, we trained a binary classifier computer vision model to predict whether each tree crop depicted a dead or live tree. The training dataset for the dead/live classifier was constructed by first identifying a subset of

drone-detected crowns matched to dead field trees and manually verifying the correctness of each match (i.e., that the tree appeared dead). Verified matches were labeled "dead" and all remaining trees were labeled "live". The final dataset comprised a training set of 12,110 images (4,074 dead, 8,036 live) from 209 trees, a validation set of 4,050 images (1,355 dead, 2,695 live) from 60 trees, and a test set of 1,751 images (558 dead, 1,193 live) from 30 trees.

We fine-tuned an EfficientNetV2 model (Tan & Le, 2021) using the MMPretrain framework (MMPreTrain Contributors, 2023), initialized with ImageNet-21k pretrained weights. The model was trained using AdamW optimizer with a linear warm-up over the first 3 epochs followed by cosine annealing, and a batch size of 128. Data augmentation included random resized cropping, horizontal flipping, random rotation, and brightness, contrast, and gamma adjustments.

We then used the trained model to run inference on all tree crops and aggregated predictions by majority voting at the tree level to produce a single dead/live label per tree. Only trees classified as live were retained in the dataset prepared for multi-view species prediction model training and evaluation.

Data preparation for computer vision model training

We constrained our modeling experiments to the eight most common species classes among our paired drone-ground tree dataset: *Abies concolor* (ABCO), *Abies magnifica* (ABMA), *Calocedrus decurrens* (CADE27), *Notholithocarpus densiflorus* (NODE3), *Pinus contorta* (PICO), *Pinus lambertiana* (PILA), *Pinus ponderosa* (PIPO) and *Pinus jeffreyi* (PIJE) together as PIPJ, and *Pseudotsuga menziesii* (PSME).

To train the computer vision model for the image classification task, we divided the dataset into training, validation and test subsets. When creating these splits, we ensured that all trees from any given plot were assigned exclusively to either the training or validation set to prevent data leakage and any bias caused due to spatial autocorrelation (Ploton et al., 2020; Roberts et al., 2017). Since our region of study had an imbalance in distribution across species classes, it was also important to maintain a similar species class frequency across the subsets to ensure each subset reflects the overall class distribution of the dataset.In standard ML practices, stratification is typically applied at the sample level. In our case, since data splits are defined at the plot level, we adopted an optimization-based stratified splitting strategy to achieve relatively consistent class frequency between the train and test sets:

a. Species composition matrix:
   We first constructed a species quantity matrix where each row corresponds to a plot and each column corresponds to a tree species class. For each plot-species pair, we counted the number of individual trees and entered this value into the matrix. This matrix captures the detailed species composition of every ground plot and serves as a foundation for the split balancing.

If N_{ij} represents the number of trees of species j in plot i, then the resulting matrix M = [N_{ij}] summarizes all species distributions across plots. To account for plots surveyed by multiple drone mission pairs, each plot's tree counts were weighted by the number of flight pairs assigned to it, so that plots covered by multiple drone flight pairs contribute proportionally more to the balance score.

b. Optimization-based split selection:
To identify a split that minimizes the difference in species composition between the training and validation datasets, we performed a stochastic search. The algorithm sampled 100,000 candidate splits, each with approximately a random 20% of plots assigned to the validation set. Then for each candidate split:

1. The total number of trees per species was calculated separately for the training and validation subsets.
2. The total values were normalized into proportions to yield per-species relative abundances in each subset.
3. The balance score was computed as the sum of squared differences between the two proportion vectors:
$$S = \Sigma \left(p_{train,\,j} - p_{val,j}\right)^2$$

where $p_{train,\,j}$ and $p_{val,j}$ denote the relative abundances of species j in the training and validation sets respectively. The algorithm selects the split with the minimum balance score S, corresponding to the most similar species composition across subsets.

For model testing, we used the set of approximately 20% of Open Forest Observatory ground plots and drone mission pairs designated by the Open Forest Observatory as testing-only datasets. (These datasets were not used for training.) The testing-only datasets were selected via stratified random sampling with constraints aimed at selecting a set of plots with a distribution of forest structure, species composition, and abiotic conditions representative of the full catalog. Some drone missions spatially overlap with others, causing a potential for data leakage between train and test sets if the missions are not kept together. To account for overlaps, all missions in a group of overlapping missions–along with any ground plots they overlapped–were kept together in either the test set or training set. This approach ensures that all images/views for a tree and all trees from a plot belong to the same split, while maintaining proportionate class distribution across train and validation splits. The 133 ground_plot–HN–LO sets were split into 86 training plots with 5,741 trees (from 70 unique plots), 21 validation plots with 1,601 trees (from 18 unique plots), and 26 test plots with 1,896 trees (from 19 unique plots) (Fig. 2).

The composite test set contained 1,896 trees in total, of which 1,894 had at least one oblique view and 1,311 had at least one nadir view. The smaller nadir subset reflects the higher rate of exclusion due to the minimum chip size threshold, while oblique exclusions were negligible (2 trees).

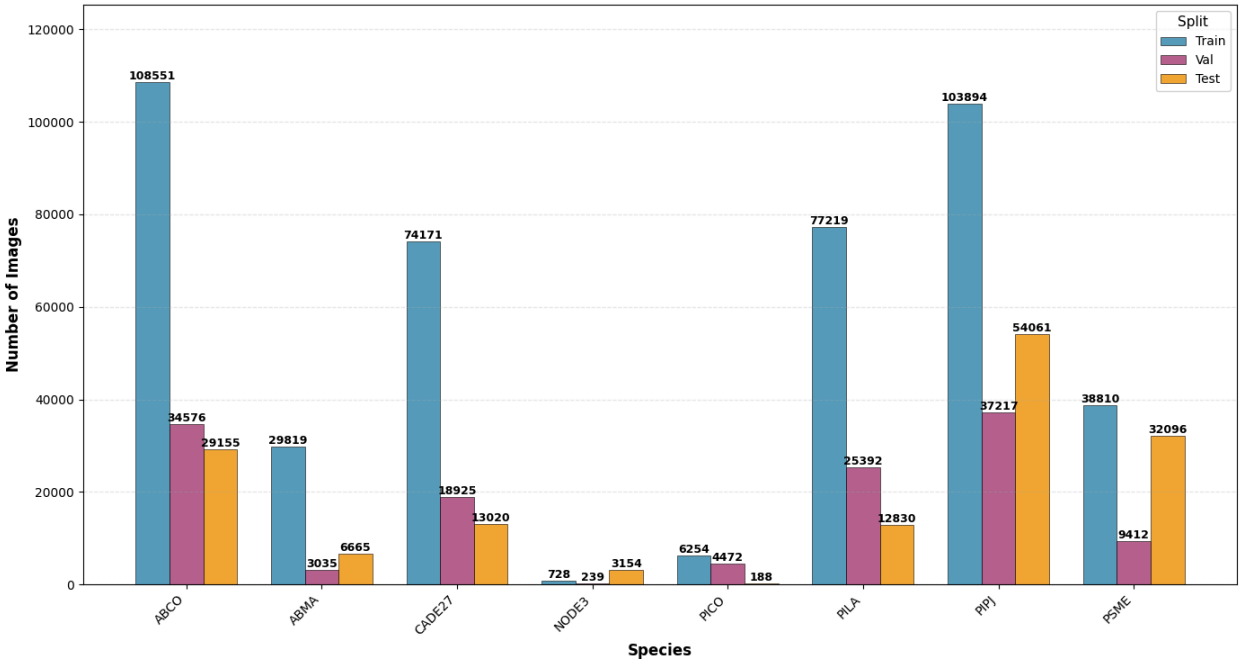

Figure 2: Number of images across train, validation, and test splits for the eight tree species classes.

Computer vision model training and evaluation

To perform tree species prediction, we trained an image classification model. We used the EfficientNetV2 architecture implemented in the open-source MMPretrain image classification framework. The model was initialized with pretrained ImageNet weights and performed transfer learning to fine-tune the model on the prepared tree species dataset.

Model training was conducted for 100 epochs using the AdamW optimizer with learning rate 2.54E-04, weight decay as 0.001, and gradient clipping (max norm set to 5.0). These hyperparameters were chosen from the best performing experiment after hyperparameter optimization on the validation set. A linear warm-up was applied over the first ten epochs, followed by a cosine annealing learning-rate schedule. The training, validation, and test sets each used a batch size of 256. To address class imbalance, we used focal loss ($\alpha = 0.25$, $\gamma = 2.0$) and MixUp and CutMix augmentations during training to improve model robustness. Other data augmentation included randomly cropping images to a region covering 60-100% of the original image area before resizing to 224x224 pixels, along with horizontal flipping and brightness and contrast adjustments. Images were then normalized using standard ImageNet mean and standard-deviation values. Model performance was monitored after each epoch using top-1 accuracy metrics on the validation set, and the best checkpoint was selected automatically based on highest validation tree-level performance.

To evaluate the model performance, we used three metrics: precision, recall, and F1 score. For a given class, precision is defined as the fraction of predictions assigned to that class that were correct, while recall is the fraction of true instances of that class that were correctly identified by the model. The F1 score is the harmonic mean of precision and recall, and acts as a single metric that balances both. Since our dataset had class-imbalance, we report both macro-averaged metrics (giving equal weight to each class) and overall accuracy, which is the fraction of all instances correctly classified.

We conducted evaluation at two levels of data aggregation: image-level and tree-level. At the image level, each individual view is treated as an independent prediction, and accuracy is computed across all views in the test set. This measures how well the model performs on any given view of a tree, but does not account for the fact that multiple images of the same tree share the same ground-truth label. At the tree level, predictions from all images belonging to the same tree are aggregated into a single prediction through majority voting across all views of a tree, and the class with the most votes is taken as the final prediction. Tree-level evaluation thus reflects the model's practical utility, where the end goal is a single species label per tree rather than per image.

Modeling experiments

Species class aggregation

We performed a number of experiments involving model training and/or inference with alternative data subsets and class aggregation approaches. First, we evaluated whether using coarser species classes–which merge species that are related taxonomically, morphologically, and functionally and may be suitable for many management applications–improves model accuracy. We merged the 8 base classes to 4 coarse classes by lumping the two fir (*Abies*) species together with Douglas-fir (*Pseudotsuga menziesii*) as “FIR” and by lumping the three pine (*Pinus*) species as “PI”. We relabeled the one broadleaf (“hardwood”) species (tanoak, *Notholithocarpus densiflorus*) as “HW” and the one species of the family Cupressaceae (incense cedar or *Calodedrus decurrens) as “*CUPR”. We evaluated merging the classes both post-training and post-inference (i.e., by predicting base-level classes and merging them after inference prior to evaluation), and pre-training and pre-inference (i.e., by training an alternative model using the coarse classes and evaluating its coarse-class predictions directly). The coarse-class model used the same set of training and evaluation images; only the labels differed.

For coarse-level classes with one-to-one mappings (CUPR, HW), the granular model's prediction was directly mapped. For aggregated classes comprising multiple granular species (FIR, PI), the aggregated class probabilities were computed as the maximum predicted probability across the constituent classes, and the final prediction for the image was the class with the highest probability.

Image perspective: Orthomosaic vs. nadir vs. oblique vs. composite

To compare and understand the improvements from the composite multi-view species classification approach, we trained an EfficientNetV2 model on the cropped tree images from the orthomosaics alone. In the orthomosaic dataset, each tree has exactly one top-down image, whereas each tree in the multi-view dataset includes multiple images (avg. 74 views) from different perspectives. We also compared these models to the performance of models trained with nadir-only and oblique-only images.

Subregion-specific training

To understand whether subregion-specific training improves model performance when predicting tree species at sites within that subregion, we trained and evaluated an alternative EfficientNetV2 using data only from sites within the North Yuba River watershed (28 of 86 training plots and 16 of 26 test plots). For fair comparison against the all-sites model, we additionally evaluated the all-sites model on the same subsetted (11 plots) test set.

Results

EfficientNetV2 performance on baseline test dataset

The EfficientNetV2 model fine-tuned and evaluated using the baseline dataset (full-region, fine-grained species classes, composite multi-view imagery) achieved 85.81% overall accuracy from image-level evaluation on the unseen test dataset across the eight species classes (Table 1). Recall for all classes improved substantially (+2-26% depending on the species) when predictions were aggregated at the tree level. Precision for most classes also generally improved (+6-57%), with the only exceptions being PIPJ (-3.4%), PSME (-8.9%) and ABCO (-0.64%). The majority classes ABCO, PILA, PIPJ, and PSME achieved tree-level recall values greater than the overall 85.81%, while the minority classes achieved lower values. Notably, the minority class PICO, which had only 256 trees in the training set and 7 trees in the test set, performed worst, with 46% recall at the image level. However, when aggregated at the tree level, the class was correctly predicted for 5 out of the total 7 PICO trees, improving recall to 71.43%.

Dead/Live classification EfficientNetV2 model – The best performing checkpoint was selected based on validation performance, and this model achieved a tree-level macro-averaged recall of 97.6% on the held-out test set.

Table 1: Baseline EfficientNetV2 model results for eight species classes evaluated at image level and tree level on the test set.

Species Class	Total Images	Total Trees	Avg views / tree	Recall (image level)	Precision (image level)	Recall (tree level)	Precision (tree level)
ABCO	29,155	361	94	85.55%	82.16%	89.20%	81.52%
ABMA	6,665	186	39	66.62%	62.05%	74.19%	89.03%
CADE27	13,020	308	46	75.75%	78.85%	81.17%	86.51%
NODE3	3,154	111	31	34.15%	89.97%	40.54%	95.74%
PICO	188	7	27	45.74%	14.14%	71.43%	71.43%
PILA	12,830	115	116	84.51%	81.72%	86.09%	94.29%
PIPJ	54,061	458	129	90.94%	96.08%	94.10%	92.69%
PSME	32,096	350	106	91.30%	83.96%	92.86%	75.06%
Macro Average				71.82%	73.62%	78.70%	85.78%
Overall Accuracy				85.81%		85.18%	

The most common class-level confusions were of ABCO and ABMA as one another and of CADE27 and NODE3 as PSME (Fig. 3a).

(a) Baseline model at image level

(b) Baseline model at tree level

(c) Coarse-class model at tree level

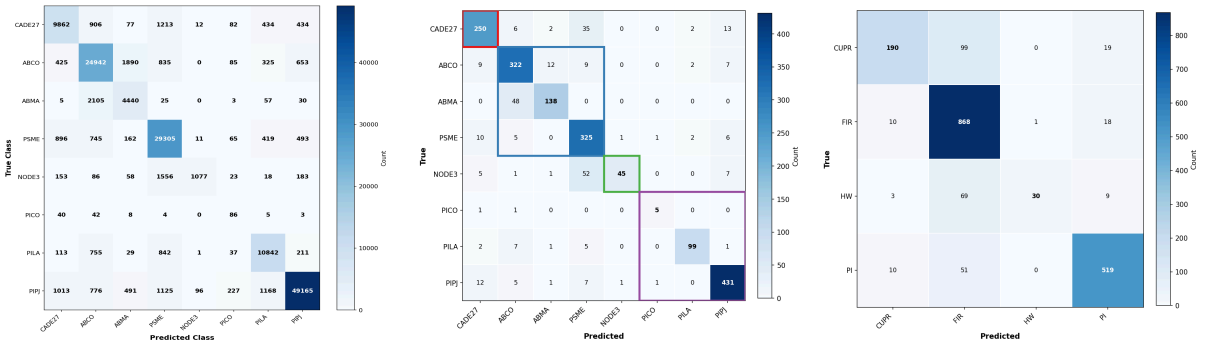

Figure 3: Confusion matrices of baseline model at (a) image level and (b) tree level evaluation results on test dataset. Colored borders in (b) represent base-level class aggregation as follows – (i) red groups *Cupressaceae* (CUPR) class, (ii) blue groups *Abies* & Douglas-fir (FIR) classes, (iii) green groups hardwood (HW) class, and (iii) purple groups *Pinus* (PI) classes. Confusion matrix (c) shows tree level evaluation results from a model trained on these coarsely defined classes.

We found that tree-level accuracy increased with the number of views, rising from ~0.68 for trees with fewer than 5 views to ~0.90 for trees with more than 200 views (Fig. 4).

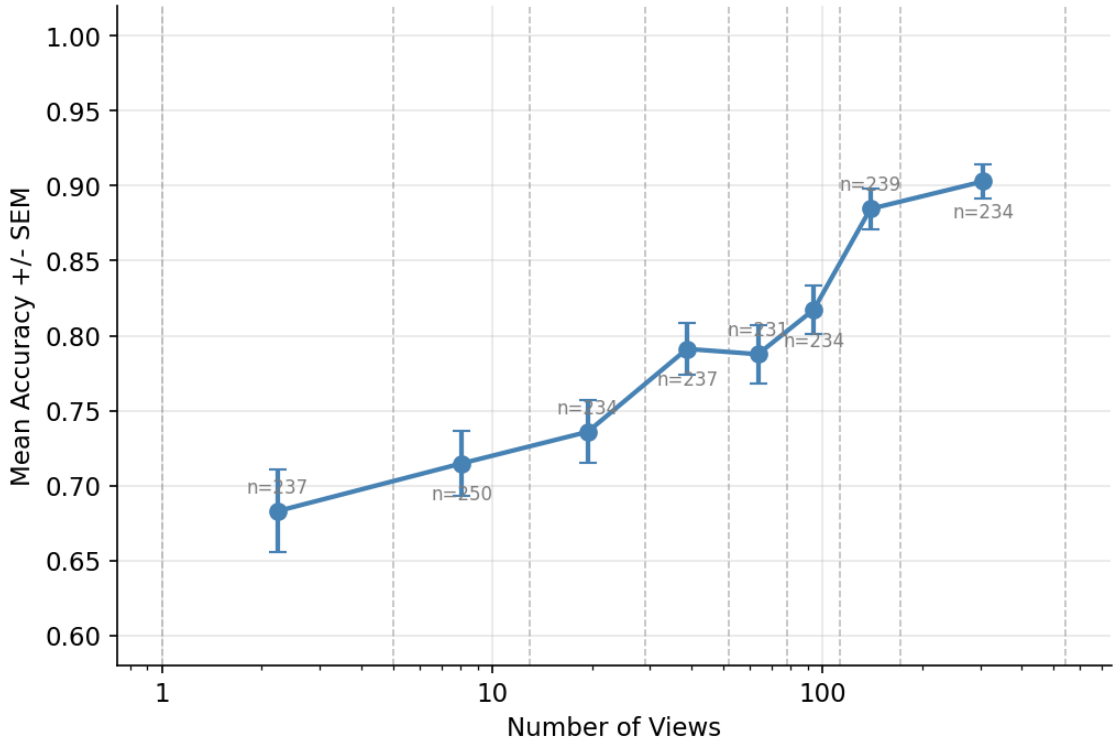

Figure 4: Effect of number of views on tree-level classification accuracy. Points indicate mean accuracy (+- SE) for each of 8 different bins of per-tree image count, with bin boundaries determined by evenly-spaced quantiles in log space. For each bin, the mean fraction of correctly classified views across trees was computed, with error bars representing the standard error of the mean (SEM).

Performance comparison: Granular vs. coarse species classes

Table 2: Recall (Coarse-class model) summarizes coarse-class model's recall on the test set. Recall (Granular preds mapped to coarse classes) shows recall values after mapping granular species classes to coarse-level classes.

Species Class	Total trees	Recall (Coarse-class model)	Precision (Coarse-class model)	Recall (Granular preds mapped to coarse classes)	Precision (Granular preds mapped to coarse classes)
CUPR	308	61.69%	89.20%	80.52%	89.21%
FIR	897	96.77%	79.85%	96.21%	86.73%
HW	111	27.03%	96.77%	37.84%	95.45%
PI	580	89.48%	91.86%	92.76%	92.92%
Macro Average		68.74%	89.42%	76.83%	91.08%
Overall Accuracy		84.75%		89.19%	

Compared to the baseline model's precision and recall values on the 8 species classes (Table 1), when the model's granular predictions were mapped to coarse classes, we saw a 4% gain in overall accuracy. Inference for coarser species classes was generally more accurate, as quantified via both recall and precision, when performed via post-inference aggregation (using the baseline model) than pre-training (using the alternative coarse-class model) (Table 2). Pre-training species class lumping particularly hurt recall for the CUPR class and, to a lesser extent, recall for the FIR class. Confusions outside of these coarse-level groupings were less common when the lumping was performed post-inference, and post-inference lumping resolved some fine-grained confusion within groups of ultimately lumped species (Fig. 3b-3c).

Performance comparison: Orthomosaic vs. nadir vs. oblique vs. composite

Table 3: Tree-level performance comparison between EfficientNetV2 trained with orthomosaic data, nadir images, oblique images, and composite multi-view images. Nadir subset excludes trees filtered by minimum chip size threshold. Recall (Nadir) % is computed only over trees with at least one retained nadir view; Recall (Nadir) % incl. missing trees treats trees lacking nadir views as incorrect predictions. The oblique test set had negligible exclusion (2 trees).

Species Class	Total trees	Recall (Ortho) %	Recall (Ortho) % only trees with nadir view(s)	Recall (Nadir) %	Recall (Nadir) % incl. missing trees	Recall (Oblique) %	Recall (Composite) %
ABCO	361	80.29%	82.18%	90.81%	71.19%	82.27%	89.20%
ABMA	186	39.25%	47.57%	65.05%	36.02%	62.90%	74.19%
CADE27	308	85.67%	85.61%	73.61%	34.42%	71.75%	81.17%
NODE3	111	5.41%	5.00%	12.50%	4.50%	54.95%	40.54%
PICO	7	57.14%	40.00%	40.00%	28.57%	85.71%	71.43%
PILA	115	72.32%	73.27%	84.62%	76.52%	84.35%	86.09%
PIPJ	458	88.21%	90.32%	94.05%	72.49%	93.45%	94.10%
PSME	350	73.82%	81.99%	90.32%	72.00%	90.23%	92.86%
Macro Average		62.76%	63.24%	68.87%	49.46%	78.55%	**79.42%**
Overall accuracy		72.61%	78.61%	84.59%	58.49%	81.36%	**85.18%**

Species prediction accuracy was generally greatly improved by using the multi-view composite (nadir and oblique images) dataset compared with the orthomosaic (single top-down view) dataset, with macro-averaged recall and overall accuracy increasing by 17% and 13%, respectively (Table 3). All species were predicted with greater accuracy (+6-35% recall) with the

exception of CADE27 (-4.5% recall). Across the three multi-view datasets (nadir-only, oblique-only, and composite), the composite model achieved the highest macro-averaged recall (79%) — 1% higher than the oblique-only dataset and 30% higher than the nadir-only dataset when counting missing trees as incorrect classifications (Table 3). The overall accuracy using the composite dataset (85%) was 4% higher than the oblique-only dataset and 27% higher than the nadir-only dataset.

Performance comparison: Yuba region specific model vs. all sites model

Table 4: Tree-level performance on the Yuba test set compared between the Yuba-specific model and the full (baseline) model.

Species Class	Total trees	Recall (Yuba model)	Precision (Yuba model)	Recall (Full model)	Precision (Full model)
ABCO	221	97.29%	54.43%	92.76%	79.77%
ABMA	185	15.14%	100.00%	74.05%	93.84%
CADE27	34	41.18%	73.68%	67.65%	65.71%
NODE3	103	79.61%	93.18%	41.75%	95.56%
PICO	1	0.00%	0.00%	0.00%	0.00%
PILA	29	72.41%	91.30%	89.66%	92.86%
PIPJ	80	91.25%	89.02%	92.50%	86.05%
PSME	111	82.88%	74.80%	88.29%	58.68%
Macro Average		59.97%	72.05%	68.33%	71.56%
Overall Accuracy		68.71%		79.31%	

When evaluated against the Yuba-specific test set, the full model outperformed the Yuba-specific model on nearly all metrics (Table 4): overall accuracy on the Yuba test set was 79.31% when using the full model, compared to 68.71% when using the Yuba-only model. The full model's macro-averaged recall exceeded that of the Yuba model by 8.4%, while macro precision was comparable between the two.

Discussion

Multi-view imagery improves species classification over orthomosaics

The multi-view model applied to the composite (nadir and oblique images) dataset achieved an accuracy improvement of more than 16% when compared to the orthomosaic model. This gain in accuracy could have resulted from a combination of three factors. First, raw drone imagery naturally captures trees from multiple perspectives, including oblique angles that expose bark

texture, branching structure, and crown morphology– features that ecologists themselves rely on for field identification but that are practically invisible in top-down canopy views (Michałowska et al., 2023; Natesan et al., 2019). Second, orthomosaics introduce undesired image artifacts (James & Bradshaw, 2020; Russell et al., 2024) and distortions during the stitching process that can degrade fine-grained visual features (Fig. 1a), whereas raw drone imagery preserves the original image quality. Lastly, the multi-view dataset provided a much larger training sample relative to the orthomosaic. Each tree yields exactly one image when cropped from an orthomosaic, while the multi-view dataset contains roughly 20-100 images per tree. As computer vision models generally benefit from larger training sets (Sun et al., 2017), some portion of the performance gap may be due to this difference. That said, the natural expansion of the dataset through multiple real-world perspectives is itself one of the central advantages of the multi-view approach. Unlike synthetic augmentation techniques such as rotation or brightness jittering, each additional view captures naturally distinct visual information about the tree under different lighting conditions, angles, and background information, which makes it a practical and meaningful benefit. Additionally, the availability of multiple views per tree enables prediction aggregation through majority voting, an ensemble mechanism that reduces the influence of noisy or ambiguous individual views on the final species assignment.

At the species level, only one of our eight species, CADE27 (incense cedar) was predicted more poorly via the multi-view model vs. the orthomosaic model. This species might appear more unique from a top-down perspective, whereas in side views, it can be confused with other species, apparently particularly PSME (Douglas-fir) (Fig. 1b). As an alternative explanation, the segmentation of drone-detected trees and linking of field-surveyed species labels to drone-detected trees undoubtedly include some error, such as tree crown segmentation including multiple true crowns of different species that are assigned a single species label. Perhaps from a top-down (orthomosaic) view, such clusters are more readily identified as containing incense cedar, whereas some oblique views may primarily consist of other species. The possibility of incorrect species labeling may also explain the overall poor recall of species class NODE3 (tanoak), the only broadleaf tree species in the dataset. It is a generally shorter tree that tends to be intermixed particularly with species PSME (Douglas-fir), and incorrect species class assignment in the training data might explain the high frequency of model confusion of NODE3 with PSME (Fig. 1b) and very low recall of the orthomosaic model and the multi-view model using nadir imagery. In contrast with the CADE27 confusion with PSME, the NODE3 confusion with PSME is largely resolved by including oblique views, perhaps because in the case of this species, the defining characteristics of NODE3 are more apparent in side views.

The gap between nadir recall computed only over trees with at least one nadir view vs. nadir recall including missing trees (Table 3) — dropping from 74% to 34% for CADE27 and 65% to 36% for ABMA — largely reflects the fact that we treated trees with no views as incorrectly predicted. This large drop highlights that relying only on high-altitude nadir images, despite their superiority for reconstructing 3D structure and detecting trees (Young et al. 2022), can fail to yield sufficient high-resolution views of trees for computer vision-based species prediction. This observation–in combination with the overall greater accuracy of the composite dataset over

either oblique or nadir datasets alone–helps to justify the additional time involved in performing a secondary low-altitude oblique imaging drone mission in addition to the high-altitude oblique mission. When excluding trees with no nadir views (rather than treating them as incorrectly predicted), some species exhibited better recall relative to using oblique views, but this likely reflects the fact that the trees with nadir reviews remaining following low-resolution view filtering were larger, more prominent trees with more clearly visible distinguishing features.

Tree-level prediction aggregation improves classification

Aggregating image-level predictions to the tree level through majority voting consistently improved recall and precision across nearly all species classes. This result is expected since when a tree is observed across multiple independent views, classification errors on individual images which can be caused by partial occlusion, variable lighting, or background noise tend to cancel out (Russell et al., 2024). This interpretation is further supported by the strong increase in prediction accuracy as view count increased (Fig. 4), though this pattern may also be produced if larger trees, with more distinctive visual structure, have more views and more informative views because they are less occluded by their neighbors. Nonetheless, the accuracy gains from aggregating multiple views were most pronounced for rare classes like PICO, which showed improved recall (+25.69%) and precision (+57.29%) after aggregation at the tree level, suggesting that the multi-view approach is particularly valuable for minority classes for which model confidence in any given image is low and predictions are unstable, or where individual image quality is variable.

Hierarchical mapping from granular predictions outperforms direct coarse-class training

When predictions from the granular species model were mapped to the four coarse-level classes, the resulting accuracy (89.19%) exceeded that of a model trained directly on coarse-level classes (84.75%). This pattern is consistent with existing reports that training a classifier using low-level species and aggregating predictions to broader groups has been shown to outperform training directly on those groups (Silla & Freitas, 2011). Fine-grained supervision can improve coarse-level classification accuracy by encouraging the network to learn more discriminative features and avoids confusing a model by telling it that multiple distinctly-appearing trees all belong to a single class (Chen et al., 2019). This stronger performance from post-inference aggregation is convenient for practical application, as (a) it means that only a single model is needed, regardless of whether the end user wishes for fine- or coarse-grained taxonomic predictions, and (b) if a given application requires only coarse taxonomy, it can enjoy higher accuracy. Many western U.S. forest ecology and management applications, for example, focus especially on the distinction between pines (*Pinus* spp.), non-pine conifers, and broadleaf tree species (Young et al., 2020; Brodie et al., 2023; Zald et al., 2024), which represent aggregations of species-level classes.

Geographic diversity in training data improves predictive accuracy within a subregion

The results of the subregion-specific experiment suggest that training on a larger, geographically diverse dataset benefits model predictive accuracy even when predicting species within a specific subregion. The full model outperformed the Yuba-only model on the Yuba test set across nearly all metrics, despite the Yuba-only model having been trained exclusively on data from that region. By encountering each species across a wider range of site conditions, the full model likely learned more robust and species-discriminative features that transferred well within the subregion. The per-species results show that Yuba-only model performed especially poorly on ABMA and CADE27 (Table 4), likely because these species were underrepresented in the Yuba-only training data. However, the Yuba-only model achieved higher recall for NODE3, a species uniquely abundant in the Yuba subregion relative to the full model domain, suggesting that in cases where important species are well represented locally but rare globally, regional training can be beneficial.

Synthesis, practical deployment, and future pursuits

Our workflow that combines composite multi-view raw drone imagery and computer vision to accurately identify individual tree species across a geospatially diverse set of yellow pine & mixed-conifer forests in the western U.S. By leveraging the overlapping nature of raw drone images, our approach generates tens to hundreds of views per tree combining nadir and oblique views. The predictions are aggregated across these images to produce a single tree-level species prediction, achieving substantial improvements over traditional single-view orthomosaic methods.

When selecting a computer vision modeling framework for tree species prediction, analysts have the choice between classification models, which apply a label to an entire image, and segmentation models, which label each pixel of an image with potentially differing labels. Previous multiview tree species classification work has used a segmentation framework, labeling all pixels of a given drone image (generally containing multiple trees) with species classes. This approach limits the efficiency and practicality of deployment because it requires rendering of all tree instances for each new model run, which can be computationally expensive. Unlike segmentation-based approaches, the classification framework presented here decouples the rendering and inference steps. Tree instance IDs are rendered once and the resulting crops can be reused across multiple independent model runs. For example, predicting both species and live/dead status, subsampled to different view subsets, or class aggregations, without repeating the rendering step. These various advantages make classification models much more efficient.

Although the multi-view approach clearly improves predictive accuracy, the relative contributions of increased view diversity and increased training data volume remain unclear. A useful direction for future work would be a controlled comparison, such as training an orthomosaic model augmented to match the multi-view dataset size, to better understand this. Additionally, while our results demonstrate that aggregating predictions across multiple views improves species classification, it remains unknown which specific visual features and view perspectives drive correct predictions for each species. Applying explainable AI techniques such as

Grad-CAM to identify the most informative features and view angles across the different species could directly inform data collection strategies–for example, by identifying optimal camera angles or altitudes for a region dominated by a particular species–and guide model improvements by revealing whether predictions rely on genuine morphological features or on uninformative features such as background pixels or lighting artifacts. This could motivate targeted improvements to data collection or crown masking strategies. While there are many promising opportunities for future gains in accuracy and deployment efficiency, our workflow already provides a scalable, deployment-ready approach for automating accurate, geospatial tree-level species predictions across diverse forest landscapes, with direct applications including forest inventory, carbon monitoring, and wildfire risk assessment and mitigation.

Data Availability

Data is being prepared for release with the next preprint version. The codebase and documentation for reproducing all our experiments is available at https://github.com/open-forest-observatory/tree-species-prediction

Acknowledgements

We are grateful for the extensive efforts of the numerous groups and individuals who contributed drone data and ground reference data that they collected to the Open Forest Observatory Data Catalog. We thank B. Pardi for sharing feedback on early versions of this workflow. This work was supported by the National Science Foundation (grant #2152671) and the California Department of Forestry and Fire Protection Forest Health Research Program via the California Climate Investments program (grant #8GG24804). This work used Jetstream2 cloud computing resources at Indiana University through allocation BIO220124 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296.

References

Brodie, E. G., Knapp, E. E., Latimer, A. M., Safford, H. D., Vossmer, M., & Bisbing, S. M. (2023). The century-long shadow of fire exclusion: Historical data reveal early and lasting effects of fire regime change on contemporary forest composition. *Forest Ecology and Management*, *539*, 121011. https://doi.org/10.1016/j.foreco.2023.121011

Chen, Z., Ding, R., Chin, T.-W., & Marculescu, D. (2019). *Understanding the Impact of Label Granularity on CNN-based Image Classification* (arXiv:1901.07012). arXiv. https://doi.org/10.48550/arXiv.1901.07012

Hsieh, P.-F., & Lee, L.-C. (2000). Effect of spatial resolution on classification error in remote sensing. *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120)*, *1*, 171–173 vol.1. https://doi.org/10.1109/IGARSS.2000.860458

Huang, Y., Ou, B., Meng, K., Yang, B., Carpenter, J., Jung, J., & Fei, S. (2024). Tree Species Classification from UAV Canopy Images with Deep Learning Models. *Remote Sensing*, *16*(20), 3836. https://doi.org/10.3390/rs16203836

Immitzer, M., Atzberger, C., & Koukal, T. (2012). Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. *Remote Sensing*, *4*(9), 2661–2693. https://doi.org/10.3390/rs4092661

James, K., & Bradshaw, K. (2020). Detecting plant species in the field with deep learning and drone technology. *Methods in Ecology and Evolution*, *11*(11), 1509–1519. https://doi.org/10.1111/2041-210X.13473

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*. https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Liu, T., & Abd-Elrahman, A. (2018). An Object-Based Image Analysis Method for Enhancing Classification of Land Covers Using Fully Convolutional Networks and Multi-View

Images of Small Unmanned Aerial System. *Remote Sensing*, *10*(3), 457. https://doi.org/10.3390/rs10030457

Meyer, F., & Beucher, S. (1990). Morphological segmentation. *Journal of Visual Communication and Image Representation*, *1*(1), 21–46. https://doi.org/10.1016/1047-3203(90)90014-M

Michałowska, M., Rapiński, J., & Janicka, J. (2023). Tree species classification on images from airborne mobile mapping using ML.NET. *European Journal of Remote Sensing*, *56*(1), 2271651. https://doi.org/10.1080/22797254.2023.2271651

MMPreTrain Contributors. (2023). *OpenMMLab's Pre-training Toolbox and Benchmark* (Version 0.15.0) [Python]. https://github.com/open-mmlab/mmpretrain (Original work published 2020)

Natesan, S., Armenakis, C., & Vepakomma, U. (2019). RESNET-BASED TREE SPECIES CLASSIFICATION USING UAV IMAGES. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2-W13*, 475–481. ISPRS Geospatial Week 2019 (Volume XLII-2/W13) - 10–14 June 2019, Enschede, The Netherlands. https://doi.org/10.5194/isprs-archives-XLII-2-W13-475-2019

Observatory, O. F. (2026). *Open Forest Observatory Public Forest Inevnentory Database*. https://doi.org/10.5281/zenodo.20835533

Ohmann, J., & Gregory, M. (2002). Predictive mapping of forest composition and structure with direct gradient analysis and nearest- neighbor imputation in coastal Oregon, U.S.A. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere - CAN J FOREST RES*, *32*, 725–741. https://doi.org/10.1139/x02-011

Onishi, M., & Ise, T. (2021). Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific Reports*, *11*(1), 903. https://doi.org/10.1038/s41598-020-79653-9

Pallavoor, A., Russell, D., Chen, M., & Young, D. (2025). *open-forest-observatory/tree-detection-framework: Release v0.1.1* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.14915261

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, *11*(1), 4540. https://doi.org/10.1038/s41467-020-18321-y

Popescu, S. C., & Wynne, R. H. (2004). Seeing the Trees in the Forest. *Photogrammetric Engineering & Remote Sensing*, *70*(5), 589–604. https://doi.org/10.14358/PERS.70.5.589

Raczko, E., & Zagajewski, B. (2017). Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. *European Journal of Remote Sensing*, *50*(1), 144–154. https://doi.org/10.1080/22797254.2017.1299557

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., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Russell, D., Pallavoor, A., Bucciarelli, G., Latimer, A., & Young, D. J. N. (2025). *Mapping California woodland-chaparral ecosystems following wildfire with diverse drone images and computer vision*. https://ecoevorxiv.org/repository/view/10435/

Russell, D., Sidhu, A., Prabhu, R., Young, D., & Addisherla, T. (2025). *open-forest-observatory/geograypher: V0.3.0* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.15539928

Russell, D., Weinstein, B., Wettergreen, D., & Young, D. (2024). *Classifying geospatial objects from multiview aerial imagery using semantic meshes*. https://doi.org/10.48550/arXiv.2405.09544

Safford, H. D., & Stevens, J. T. (2017). *Natural range of variation for yellow pine and mixed-conifer forests in the Sierra Nevada, southern Cascades, and Modoc and Inyo National Forests, California, USA* (PSW-GTR-256; p. PSW-GTR-256). U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station. https://doi.org/10.2737/PSW-GTR-256

Santos, A. A. dos, Marcato Junior, J., Araújo, M. S., Di Martini, D. R., Tetila, E. C., Siqueira, H. L., Aoki, C., Eltner, A., Matsubara, E. T., Pistori, H., Feitosa, R. Q., Liesenberg, V., & Gonçalves, W. N. (2019). Assessment of CNN-Based Methods for Individual Tree Detection on Images Captured by RGB Cameras Attached to UAVs. *Sensors*, *19*(16), 3595. https://doi.org/10.3390/s19163595

Schadauer, T., Karel, S., Loew, M., Knieling, U., Kopecky, K., Bauerhansl, C., Berger, A., Graeber, S., & Winiwarter, L. (2024). Evaluating Tree Species Mapping: Probability Sampling Validation of Pure and Mixed Species Classes Using Convolutional Neural Networks and Sentinel-2 Time Series. *Remote Sensing*, *16*(16), 2887. https://doi.org/10.3390/rs16162887

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., & Schmidtlein, S. (2020). Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, *170*, 205–215. https://doi.org/10.1016/j.isprsjprs.2020.10.015

Shang, X., & Chisholm, L. (2014). Classification of Australian Native Forest Species Using Hyperspectral Remote Sensing and Machine-Learning Classification Algorithms. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal Of*, *7*, 2481–2489. https://doi.org/10.1109/JSTARS.2013.2282166

Silla, C. N., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. *Data Mining and Knowledge Discovery*, *22*(1), 31–72. https://doi.org/10.1007/s10618-010-0175-9

Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). *Revisiting Unreasonable Effectiveness of Data in Deep Learning Era* (arXiv:1707.02968). arXiv. https://doi.org/10.48550/arXiv.1707.02968

Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of the 38th International Conference on Machine Learning*, 10096–10106. https://proceedings.mlr.press/v139/tan21a.html

TreeMap 2016: A tree-level model of the forests of the conterminous United States circa 2016. (n.d.). Retrieved June 1, 2026, from https://www.fs.usda.gov/rds/archive/catalog/RDS-2021-0074

Weinstein, B. G., Marconi, S., Graves, S. J., Zare, A., Singh, A., Bohlman, S. A., Magee, L., Johnson, D. J., Townsend, P. A., & White, E. P. (2023). Capturing long-tailed individual tree diversity using an airborne imaging and a multi-temporal hierarchical model. *Remote Sensing in Ecology and Conservation*, *9*(5), 656–670. https://doi.org/10.1002/rse2.335

Weinstein, B. G., Marconi, S., Zare, A., Bohlman, S. A., Singh, A., Graves, S. J., Magee, L., Johnson, D. J., Record, S., Rubio, V. E., Swenson, N. G., Townsend, P., Veblen, T. T., Andrus, R. A., & White, E. P. (2024). Individual canopy tree species maps for the National Ecological Observatory Network. *PLOS Biology*, *22*(7), e3002700. https://doi.org/10.1371/journal.pbio.3002700

Young, D. J. N., Koontz, M. J., & Weeks, J. (2022). Optimizing aerial imagery collection and processing parameters for drone-based individual tree mapping in structurally complex conifer forests. *Methods in Ecology and Evolution*, *13*(7), 1447–1463. https://doi.org/10.1111/2041-210X.13860

Young, D., Mandel, A. I., Russell, D., Alexander, MallikaNocco, & Hereñú, D. (2024). *open-forest-observatory/automate-metashape: V0.3*. https://doi.org/10.5281/zenodo.11193943

Young, D. J. N., Meyer, M., Estes, B., Gross, S., Wuenschel, A., Restaino, C., & Safford, H. D. (2020). Forest recovery following extreme drought in California, USA: Natural patterns and effects of pre-drought management. *Ecological Applications*, *30*(1), e02002. https://doi.org/10.1002/eap.2002

Zald, H. S. J., May, C. J., Gray, A. N., North, M. P., & Hurteau, M. D. (2024). Thinning and prescribed burning increase shade-tolerant conifer regeneration in a fire excluded mixed-conifer forest. *Forest Ecology and Management*, *551*, 121531. https://doi.org/10.1016/j.foreco.2023.121531

Zeiler, M. D., & Fergus, R. (2013). *Visualizing and Understanding Convolutional Networks* (arXiv:1311.2901). arXiv. https://doi.org/10.48550/arXiv.1311.2901