Agroforestree: A Multimodal Dataset for Individual Tree Detection and Species Identification

Siddharth Sachdeva **Sidharth Tadeparti** **Hugo Ingelsson** **Isabel Lopez**
Aaditya Nalawade Aadesh Salecha Chandrashekhar Biradar David Lobell

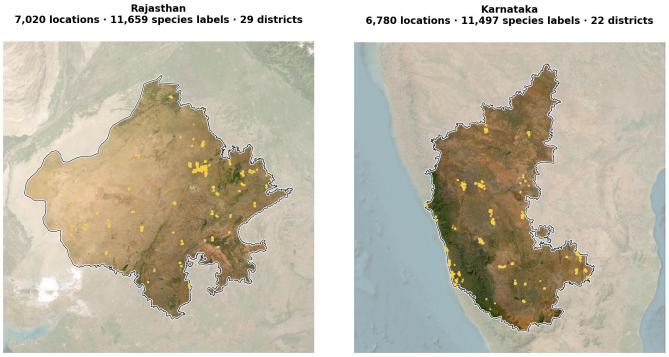
Abstract

Accurate tree species maps are needed for measuring and advancing carbon, biodiversity, and food-security outcomes in agroforestry, but labeled species data do not exist at scale. Data are particularly sparse in tropical countries, such as India, which have the greatest potential for agroforestry expansion. We develop a cross-view geolocalization pipeline, called NUTMGS, that identifies trees in high-resolution satellite imagery and links detected georeferenced tree bounding boxes to street-view images using GPS metadata and ray tracing, linking each individual tree across five co-registered sensor modalities. Applying NUTMGS yields **Agroforestree**, a training and evaluation dataset spanning three tasks — individual tree detection from satellite, species identification from street-view imagery, and species identification from remote sensing. To our knowledge, **Agroforestree** is the first expert-labeled individual-tree dataset to align each tree across five co-registered sensor modalities. It contains **353,644 expert-labeled tree crowns** across 20 Indian states and **23,156 expert species-labeled trees** in Rajasthan and Karnataka. We then benchmark whether foundation models significantly improve tree species mapping in this low-resource setting. They do not: vision foundation models do not outperform Faster R-CNN or Mask R-CNN for tree detection, BIOCLIP barely exceeds ResNet-50 for street-view species identification accuracy, and AlphaEarth and TESSERA show similarly marginal gains over MLP and XGBoost baselines for satellite species classification. These results suggest that foundation models, despite strong reported performance on existing benchmarks, need to be evaluated on the underserved regions and species where labeled data are scarcest and the scientific need is greatest.

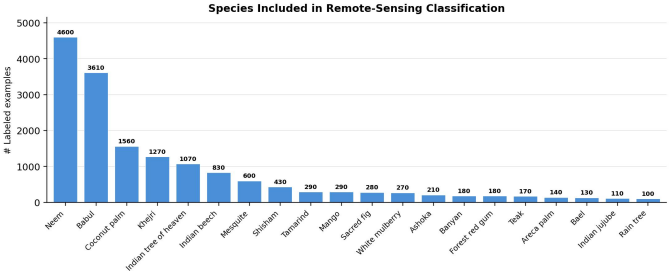
1 Introduction

Agroforestry — trees grown with crops or livestock in the same landscape — covers at least one billion hectares globally [Zomer et al., 2016], occurs predominantly in developing-country smallholder systems, and represents over 5 Gt CO_2/yr in mitigation potential [Roe et al., 2019, Cardinael et al., 2021] alongside benefits for food security and biodiversity [Nair and Garrity, 2012]. Unlike soil carbon and other nature-based solutions, tree carbon accumulation is directly measurable via satellite remote sensing, but only if tree species are known, as species-specific wood density varies tenfold and determines carbon storage [Chave et al., 2009]. Knowledge of species composition would also assist many other goals, including supply chain development and assessments of nutrition outcomes. Crop type maps have transformed field-crop monitoring at national and global scales [Bégué et al., 2018, Laguarta Soler et al., 2024], but existing maps, and ground-based benchmarks such as CropHarvest [Tseng et al., 2021], exist mainly for annual crops grown in monocultures. In polycultures such as agroforestry, where trees and crops grow in the same field, no large-scale labeled dataset or scalable labeling pipeline for individual-tree species identification exists.

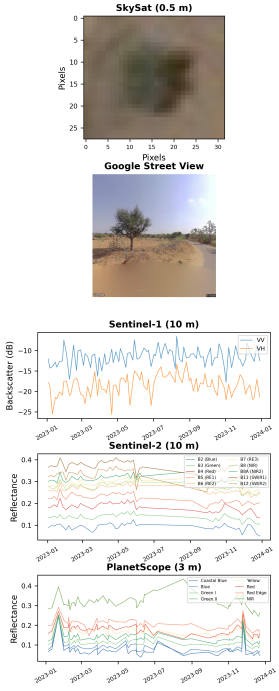
Preprint.



(a) (b)



(c) Species counts for remote-sensing classification.



(d) All five modalities for one tree.

Figure 1: **Agroforestree species release at a glance.** The 23,156 species-labeled trees span 2 Indian states (Rajasthan + Karnataka, 8 agroclimatic zones). Each tree is linked to 5 coregistered modalities at the same geolocation. Panel (d) shows all five for a single tree: 50 cm SkySat satellite crop, Google Street View panorama, and Sentinel-1 SAR, Sentinel-2 multi-spectral, and PlanetScope time series. Panel (c) shows the 20 species used in remote-sensing classification.

The primary public source of species observations for ML is iNaturalist, but its coverage of useful native trees is sparse in key agroforestry regions. Intersecting iNaturalist's research-grade observations (via GBIF, April 2026) with GlobUNT [Kindt et al., 2023] shows that **40% of useful native tree species have zero research-grade iNaturalist observations globally** and 79% have fewer than 100. The tropical gap is sharper: among ∼12,500 GlobUNT species in tropical and subtropical families, 49% have zero tropical iNaturalist observations and 91% have fewer than 100. Foundation models pretrained on iNaturalist inherit this gap. Existing alternatives do not scale to rural agroforestry: field inventories cost roughly $50–$100 per tree [Fassnacht et al., 2016], while municipal-inventory scraping requires city tree censuses and lost official Google access in May 2025 [Beery et al., 2022, Google Research, 2025]. Agroforestree fills this gap with scalable per-tree species data for useful native trees underrepresented in public ML resources.

A scalable labeling pipeline, demonstrated on agroforestry. We present **Native Useful Tree Mapping from Ground and Space (NUTMGS)**, a scalable labeling pipeline that links individual-tree detections in high-resolution satellite imagery to street-view images of the same trees, enabling expert species annotation without field inventories or municipal precursor data. The resulting labels cost approximately 2¢ per tree — roughly three orders of magnitude cheaper than field inventory — while remaining tied to single-tree geolocations and multimodal remote-sensing observations. We demonstrate the pipeline on agroforestry trees in India, producing two linked Agroforestree releases (Table 1): a continental-scale detection set (353,644 crowns across 20 Indian states) and a multimodal species subset (23,156 expert species-labeled trees in Rajasthan and Karnataka, each linked to five coregistered modalities). The detection annotations train the detector that powers the species pipeline, and the species subset supports the central evaluation: how well street-view imagery, high-resolution satellite imagery, satellite time series, and foundation models identify tree species when labels are available at individual-tree scale.

Foundation models for species identification (BIOCLIP) and for geospatial time series could produce such maps — but the per-species labeled data needed to train and validate them for understudied species and geographies does not exist.

Foundation models and Agroforestree. We evaluate three foundation-model families against task-specific baselines: vision backbones for object detection of trees in satellite imagery (DINOv3, Grounding DINO) vs. classical detectors (Faster R-CNN, Mask R-CNN, Plain-DETR), species-ID FMs (BIOCLIP [Stevens et al., 2024], zero-shot and LoRA-fine-tuned) vs. ImageNet-pretrained ResNet-50 for species identification in ground-based imagery, and geospatial FMs ((PRESTO [Tseng et al., 2023], TESSERA [TESSERA team, 2025], AlphaEarth [AlphaEarth team, 2025])) vs. MLP and TempCNN [Pelletier et al., 2019] for species identification in satellite imagery. Across detection, street-view species identification, and satellite species classification, foundation models improve some baselines but do not remove the need for in-domain labels. Agroforestree therefore serves both as a benchmark and as a data-generation pipeline for evaluating foundation models in the long-tailed tropical settings where current claims are least tested.

Contributions.

1. **NUTMGS, a scalable labeling pipeline** that produces individual-tree species labels across 5 coregistered modalities (GSV, 50 cm SkySat, 3 m PlanetScope, 10 m Sentinel-2, 10 m Sentinel-1) via cross-view geolocalization from satellite detections to GSV imagery. The pipeline runs wherever street-view coverage exists and enables species labels for understudied species and regions inaccessible to existing methods.
2. **Agroforestree, a multimodal dataset** for training and evaluating individual-tree detection and species identification across these modalities, far exceeding any existing benchmarks (Table 1), with 353,644 tree detections and 23,156 expert species labels. While some existing benchmarks have more labels for species, Agroforestree has more coregistered modalities per tree than any prior dataset (5 vs. 3 maximum, Table 1) and covers species unreachable by approaches that depend on municipal inventories or forest management records.
3. **Foundation-model evaluation** across the three tasks required for end-to-end species mapping: individual-tree detection, species identification from ground-level imagery, and species identification from remote sensing. We find that frontier foundation models (BIOCLIP, DINOv3, Grounding DINO, PRESTO, TESSERA, AlphaEarth) do not significantly outperform task-specific baselines on any of the three tasks, exposing a gap between FM claims on existing benchmarks and performance on underrepresented species and geographies.

Table 1: Public tree benchmarks compared on label granularity, scale, modality count, extensibility, and label source. N_{det}: individual-tree detection labels. N_{spec}: species labels (at the granularity shown in "Label unit"). "—" = not provided. n_{mod}: modalities linked per labeled unit. **Extensible?**: whether the labeling method can produce new labels for new species in new regions without requiring precursor data (municipal inventories, forest management records, plantation databases) that does not exist in rural agroforestry landscapes, where most of the world's smallholder tree cover occurs. Among species benchmarks, Agroforestree is the only one whose labeling method is extensible to such regions, and it has the highest modality count of any public tree benchmark (5, vs. 3 maximum in prior species benchmarks and 2 in Auto-Arborist).

Dataset	Label unit	Det. / Species	N_{det}	N_{spec}	n_{mod}	Extensible?	Label source
Individual-tree detection benchmarks							
NEON [Weinstein et al., 2019]	Individual	Det.	31k	—	1	✓	Manual annotation
ReforesTree [Reiersen et al., 2022]	Individual	Det.	4.6k	—	1	✓	Manual annotation
OAM-TCD [Veitch-Michaelis et al., 2024]	Individual	Det.	280k	—	1	✓	Manual annotation (aerial)
VHRTrees [Topgul et al., 2025]	Individual	Det.	26k	—	1	✓	Manual annotation
RWDS [Al-Emadi et al., 2025]	Individual	Det.	[PENDING: –]	—	1	✓	Manual annotation
Species benchmarks (individual, stand, plot, or patch level)							
TreeSatAI [Ahlswede et al., 2023]	Stand	Species	—	50k	3	✗	Forest management records
PureForest [PureForest team, 2024]	Plot	Species[§]	—	135k	2	✗	National forest inventory
Planted [Planted team, 2024]	Patch	Species[§]	—	2.2M	3*	✗	Plantation class records
Auto-Arborist [Beery et al., 2022][†]	Individual	Species	—	2.6M	2	✗	Municipal tree inventories
Agroforestree (ours)	**Individual**	**Det. + Spec.**	**354k**	**23k**	**5**	✓	**Manual (det.) + cross-view pipeline (spec.)**

[†] Officially turned down May 27, 2025 [Google Research, 2025]. [§] Plantation/monospecific stand level, not individual tree. * Planted uses 5 satellite sources aggregated at patch level.

2 Related Work

Tree monitoring benchmarks fall into two families, as summarized in Table 1.

Individual-tree detection benchmarks. NEON [Weinstein et al., 2019] provides ∼31k annotated trees across 22 US sites in aerial imagery. OAM-TCD [Veitch-Michaelis et al., 2024] contains ∼280k crowns in 10 cm aerial imagery. VHRTrees [Topgul et al., 2025] (∼26k, Türkiye) and RWDS [Al-Emadi et al., 2025] (multi-region, with geographic shift splits) use satellite VHR but at smaller scale. ReforesTree [Reiersen et al., 2022] contains ∼4.6k Ecuador drone trees. All provide detection labels only, in a single modality, without species. Agroforestree is larger than prior open individual-tree detection benchmarks (353k crowns vs. 280k in OAM-TCD) and, unlike the largest prior dataset, is derived from 50 cm satellite imagery rather than airborne imagery. This matters because satellite VHR is available at continental scale and on a repeat basis, while airborne coverage is expensive and geographically uneven.

Tree species benchmarks. TreeSatAI [Ahlswede et al., 2023] provides 50k patches with aerial + Sentinel-1/2 for 20 European tree species, labeled from Lower Saxony forest management records. Planted [Planted team, 2024] provides 2.2M patches across 41 countries and 5 satellite sources, but labels are at the plantation level. PureForest [PureForest team, 2024] provides 18 French species at the monospecific plot level. None of the above support individual-tree downstream use. Auto-Arborist [Beery et al., 2022] pairs GSV with aerial imagery for ∼2.6M trees across 23 North American cities, labeled to 344 genera, and served the urban forestry research community before being turned down by Google in May 2025 [Google Research, 2025]. Its labeling paradigm of joining existing municipal tree censuses to imagery cannot extend to regions without municipal inventories, which is the majority of the world's agroforestry landscapes. Agroforestree differs from prior species benchmarks by linking five coregistered modalities to each individual tree and by sourcing labels through a pipeline that can extend to new rural regions without field inventories, municipal censuses, forest management records, or plantation databases. The modality linkage matters because individual-tree species mapping needs both a precise tree-level label and remote-sensing views suitable for wall-to-wall prediction. Datasets with Sentinel-2 plus aerial imagery but no individual-tree ground label, such as TreeSatAI, cannot validate species at the crown level, while datasets with GSV plus aerial imagery, such as Auto-Arborist, do not connect labels to the satellite time series needed for regional mapping outside urban inventories.

Supervision sources for tree species ID. Existing methods for sourcing species labels from imagery are not sufficient for individual-tree multimodal species mapping because they do not

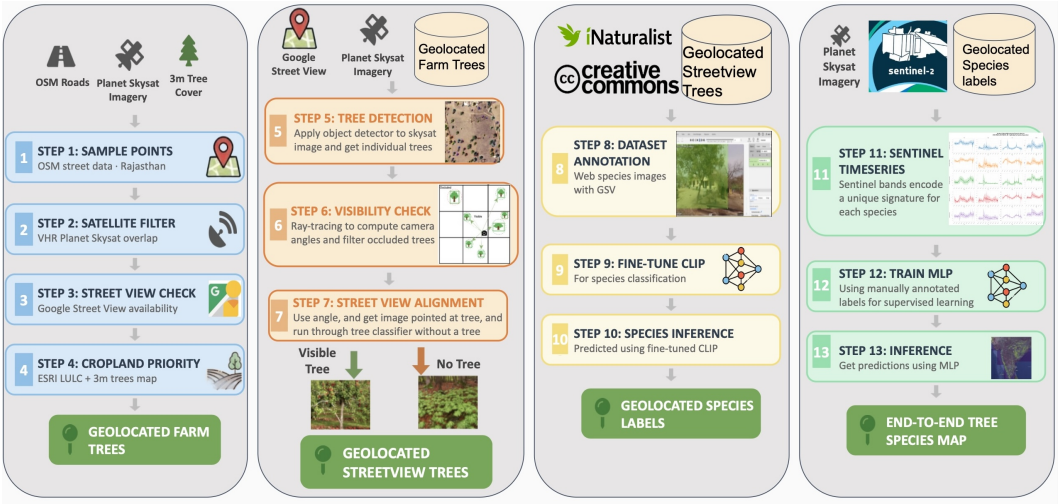


Figure 2: **The NUTMGS construction pipeline.** Starting from a trained tree detector, NUTMGS selects detected farm trees near roads, retrieves matched street-view images, obtains expert species labels, and links each labeled tree to satellite modalities for wall-to-wall mapping.

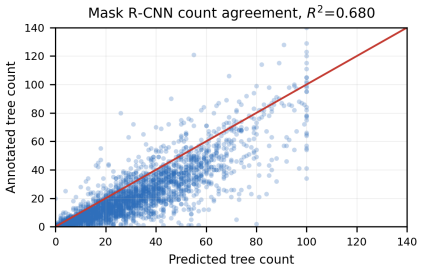
Table 2: State-balanced tree detection performance. Metrics are computed separately within each validation state and macro-averaged across the 17 states with at least 10 validation images (n=2,693 images, 61,550 annotated trees). AP_{50} uses COCO protocol. Precision and recall use detections thresholded at score 0.3 and IoU 0.5 matching. Count R^2 is squared Pearson correlation between per-image annotated and predicted tree counts. Crown RMSE is converted to meters using the 0.5 m SkySat pixel size.

Model	AP_{50}	Precision	Recall	Count R^2	Crown RMSE (m) ↓
Faster R-CNN	**63.8**	0.527	0.720	0.775	2.43
Mask R-CNN	63.4	**0.552**	0.697	0.775	2.30
Grounding DINO	41.0	0.336	**0.725**	**0.777**	2.24
Plain-DETR	41.8	0.359	0.681	0.684	2.27
Plain-DETR + DINOv3	44.5	0.372	0.684	0.721	**2.22**

provide geometrically precise tree-level labels. iNaturalist [iNaturalist, 2024] is the dominant public species-ID corpus and trains BIOCLIP [Stevens et al., 2024], but its reported coordinates identify where an observation was made, not necessarily the crown or remote-sensing pixel corresponding to the labeled tree. A well-documented North American and developed-area bias [Di Cecco et al., 2021, iNaturalist, 2019] further leaves trees in smallholder landscapes sparse. Laguarta Soler et al. [2024] showed that street-view imagery can scale crop-label collection for satellite crop mapping: a roadside image can identify the crop grown in a nearby field, and that label can train a field- or pixel-level satellite classifier. This works because many field-crop settings have large, spatially coherent fields where a small geolocation error usually preserves the crop label. The same assumption breaks in agroforestry and other polyculture systems. A single field can contain multiple tree species and crops, adjacent crowns can belong to different species, and the remote-sensing unit differs by modality, from sub-meter SkySat boxes to 10 m Sentinel pixels. Mapping species and crop composition in these systems therefore requires geometrically precise labels attached to individual plants, not only to nearby fields or street-view locations.

Foundation models for tree mapping. Vision and geospatial foundation models are increasingly evaluated on remote-sensing benchmarks, but tree-species evaluation remains narrow. DINOv3 [Siméoni et al., 2025] and Grounding DINO [Zhao et al., 2023] provide generic vision backbones for detection. Geospatial foundation models such as PRESTO [Tseng et al., 2023], TESSERA [TESSERA team, 2025], and AlphaEarth [AlphaEarth team, 2025] provide pretrained Earth-observation embeddings. TESSERA evaluates tree-species classification on TreeSatAI [Ahlswede et al., 2023], but existing evaluations remain concentrated in temperate, data-rich settings. No prior benchmark tests whether these models transfer to tropical individual-tree species in agroforestry landscapes, because geometrically precise multimodal labels for that setting did not exist. Agroforestree fills this evaluation gap.

(a) Example tree detections.



(b) Predicted versus observed tree counts.

Figure 3: **Tree detection examples and count agreement.** Detection outputs are shown alongside Mask R-CNN predicted-versus-observed tree counts from the state-balanced validation bundle before the quantitative detection benchmark.

3 The Agroforestree Benchmark

Agroforestree consists of two linked releases. The detection release provides 353,644 individual tree crowns in coregistered satellite imagery across India. NUTMGS then uses a detector trained on that release to build the species release: a subset of detected crowns with matched ground-view images, expert species labels, and links to four satellite modalities. This design makes the benchmark internally consistent: the detection data train the model used to find candidate trees, and every species example is tied back to a specific detected crown. We describe the pipeline first, then the resulting data.

3.1 NUTMGS: Native Useful Tree Mapping from Ground and Space

Figure 2 shows the full NUTMGS construction pipeline: first using the detector to identify farm trees near roads, then using geolocation and ray tracing to retrieve street-view images pointed at those trees, then obtaining species labels from the street-view images, and finally linking those labels to satellite modalities for wall-to-wall mapping.

1. **Detect farm trees near roads.** A Mask R-CNN [He et al., 2017] detector trained on the detection annotations was applied to SkySat imagery in Rajasthan and Karnataka. NUTMGS retains detected farm trees near roads as candidates for street-view matching. This step converts wall-to-wall satellite imagery into a set of individual candidate crowns with geolocations.
2. **Retrieve matched street-view images.** For each candidate crown, NUTMGS searches nearby Google Street View panoramas, uses geolocation and ray tracing to point the camera toward the tree, and filters cases where another detected crown occludes the target (Figure 2). A street-view image centered on each surviving crown is retrieved, and a linear probe on CLIP trained on 500 manually labeled images filters frames without a visible tree. Manual validation on 7,380 pairs found 83.6% of candidate GSV-image/detected-crown matches contained the target tree, with failure-mode analysis (occlusion, GPS drift, multiple-tree-in-line) in Appendix C.
3. **Annotate species in street-view imagery.** Annotators expert in Indian agroforestry species applied species labels to the retrieved GSV images (92% inter-annotator agreement on 500 double-annotated images). These species are sparsely represented in iNaturalist relative to temperate-zone tree species: each of the top-5 Rajasthan species has tens to hundreds of iNat observations globally, vs. tens of thousands for common North American and European tree species [Di Cecco et al., 2021, iNaturalist, 2019].
4. **Link species labels to remote-sensing modalities.** Each species-labeled tree anchors to its detection centroid and is linked to 5 modalities at the same geolocation: Google Street View (ground), 50 cm SkySat (VHR satellite RGB), 3 m PlanetScope (HR RGB time series), 10 m Sentinel-2 (multi-spectral time series), and 10 m Sentinel-1 (SAR time series). This places each label at sub-pixel accuracy across the PlanetScope/S2/S1 pixel grids, making the labels usable for wall-to-wall satellite species mapping rather than only patch-level image classification.

3.2 Resulting dataset

The resulting Agroforestree benchmark has two linked releases (Figure 1). The detection release provides the satellite foundation: 353,644 expert-annotated tree crowns across 20 Indian states in 50 cm Planet SkySat RGB imagery. Images were tiled into 400×400 pixel chips (200 m on a side), and annotators drew crown segmentation masks for every tree using a shadow-present requirement to distinguish trees from shrubs, following Brandt et al. [2020]. Inter-annotator agreement was 86 AP, and field measurements from 38 plots indicated 94% annotator recall. Both bounding boxes and crown segmentation masks are released.

NUTMGS uses a detector trained on the detection release as the entry point for the species benchmark. The resulting release contains 23,156 expert-labeled trees in Rajasthan and Karnataka, each linked to the same five modalities at a single detected-crown geolocation: Google Street View, SkySat, PlanetScope, Sentinel-2, and Sentinel-1. This paired structure lets the benchmark support

Table 3: Species identification across street-view and satellite/remote-sensing modalities. Values are one-vs-rest AUPRC on the held-out test set. Overall averages the 20 binary remote-sensing species; Top-5 averages the five highest-support species; Rare averages species with prevalence $<5\%$. Per-species values are in Appendix Tables 6 and 7.

Modality	Model	Overall	Top-5	Rare
GSV	BIOCLIP-2 zero-shot	0.262	0.471	0.194
	ResNet-50	0.227	0.703	0.194
	CLIP + LoRA	0.249	0.792	0.211
	BIOCLIP + LoRA	**0.335**	**0.809**	**0.302**
SkySat (50 cm)	ResNet-50 (baseline)	**0.046**	0.255	**0.032**
	CLIP + LoRA	0.044	**0.294**	0.027
	BIOCLIP + LoRA	0.034	0.261	0.018
PlanetScope ts	XGBoost	0.215	0.404	0.152
	MLP (baseline)	**0.235**	**0.439**	**0.167**
S1+S2+PlanetScope ts	XGBoost	**0.228**	**0.453**	0.153
	MLP (baseline)	0.205	0.348	**0.157**
Sentinel-2 ts	XGBoost	**0.203**	**0.411**	**0.134**
	MLP (baseline)	0.176	0.340	0.121
Sentinel-1+2 ts	XGBoost	**0.220**	**0.415**	**0.156**
	MLP (baseline)	0.178	0.321	0.130
Sentinel-1 ts	XGBoost	0.119	0.266	0.070
	MLP (baseline)	**0.165**	**0.353**	**0.102**
Geospatial FM	PRESTO (FT)	0.136	0.343	0.067
	TESSERA (FT)	**0.171**	0.372	**0.104**
	AlphaEarth (FT)	0.168	**0.401**	0.090

both individual-tree detection and per-tree species identification across ground-view and satellite modalities.

4 Evaluation

Evaluation asks four questions. First, can the detector produce reliable tree candidates across India, the entry point for NUTMGS? Second, which modality carries usable species signal for individual trees? Third, does species accuracy reliably improve with more labels, both for GSV and for remote sensing? Fourth, how do foundation-model approaches compare with task-specific baselines for detection, image classification, and geospatial embeddings? The first question is evaluated with state-balanced tree detection metrics [Lin et al., 2014]. The second and third use per-species AUPRC on a shared held-out species test set and few-shot label-scaling curves. The fourth is evaluated by comparing foundation-model rows against the strongest supervised baselines for each task.

4.1 Detection performance across India

Setup. We evaluate five detectors spanning convolutional and transformer architectures: Faster R-CNN [Ren et al., 2015], Mask R-CNN [He et al., 2017], Plain-DETR [Lin et al., 2023], Grounding DINO [Zhao et al., 2023], and frozen DINOv3 [Siméoni et al., 2025] ViT-L with a Plain-DETR head. Faster R-CNN, Mask R-CNN, and Grounding DINO are fully fine-tuned. Plain-DETR and DINOv3+Plain-DETR are trained head-only per each model's recommendation. We report India-level detection performance as a state-balanced macro-average, computing each metric separately within each state and then averaging across states so heavily sampled states do not dominate the headline result.

Main results. Table 2 reports state-balanced performance across the 17 validation states with at least 10 images. Faster R-CNN and Mask R-CNN remain the strongest AP_{50} detectors. Grounding DINO has high recall and count agreement but substantially lower precision and AP, indicating many

extra detections. Crown-size error remains near 2.2–2.4 m. Per-state AP_{50} and count-agreement results are reported in Appendix Tables 4 and 5. These per-state metrics show strong performance in most states, with a few difficult states pulling down the overall average. Lower-performing cases, such as Karnataka, are often dense, high-rainfall tree landscapes where individual crowns are difficult to separate for both annotators and models.

4.2 Species identification across modalities

Setup. We compare three image-model families on both street-view and SkySat inputs: a supervised ResNet-50 baseline, CLIP+LoRA, and BIOCLIP+LoRA. We also report zero-shot BIOCLIP-2 on GSV. For satellite time series, we evaluate XGBoost and MLP classifiers on Sentinel-1, Sentinel-2, PlanetScope, and fused time-series features. Finally, we evaluate linear probes on geospatial foundation-model embeddings from PRESTO, TESSERA, and AlphaEarth. All rows use per-species AUPRC on the shared held-out species test set. Because one-vs-rest species labels are highly imbalanced, AUPRC is more informative than accuracy: it evaluates whether a model ranks true trees above background trees, which is the first requirement for making high-precision candidate maps. Its random baseline equals class prevalence, so an AUPRC of 0.10 is random for a 10% species but 10× random for a 1% species. We therefore interpret satellite AUPRC both in absolute terms and as lift over prevalence; Appendix Table 7 reports the corresponding per-species prevalence and lift values.

Main results. Table 3 summarizes the headline species identification results across ground-level and satellite modalities. GSV is the strongest single modality, reaching 0.81 top-5 AUPRC with BIOCLIP+LoRA. For satellite-only mapping, multi-temporal features outperform single-date SkySat RGB despite coarser spatial resolution (0.41–0.45 vs. 0.29 top-5 AUPRC). PlanetScope MLP has the highest overall single-source satellite AUPRC, while S1+S2+PlanetScope XGBoost has the highest top-5 remote-sensing AUPRC. This suggests that temporal resolution and late-fusion features matter more than spatial resolution alone for many species. Absolute remote-sensing AUPRC values remain much lower than GSV, but they should be read against species prevalence. This is also why AUPRC is the right metric for the scaling loop: even before a classifier is accurate enough for final mapping, a high-AUPRC model can rank unlabeled trees so that new annotation batches are enriched for the target species. For example, S1+S2+PlanetScope XGBoost reaches 0.33 AUPRC for tamarind, whose test prevalence is only 1.6%, a 20.4× lift over random. For areca palm, it reaches 0.46 AUPRC at 0.8% prevalence, a 57.9× lift. Even for common species, where lift is necessarily smaller, neem reaches 0.58 AUPRC at 25.9% prevalence and babul reaches 0.49 AUPRC at 20.3% prevalence. Appendix Table 7 reports per-species prevalence, GSV BIOCLIP+LoRA AUPRC, S1+S2+PlanetScope XGBoost AUPRC, and lift over the random baseline.

4.3 Label efficiency and scaling with species labels

We next test whether species classification improves predictably as more positive labels are added, first for GSV and then for remote-sensing models.

GSV scaling. Figure 4 shows that street-view classification improves rapidly with modest species labels. For the 20 remote-sensing species, GSV BIOCLIP+LoRA mean AUPRC rises from 0.25 with one positive label per species to 0.55 with 100 labels, and reaches 0.63 with 1000 labels. For the highest-support species, the full BIOCLIP+LoRA model reaches 0.81 top-5 AUPRC (Table 3).

Remote-sensing scaling. This pattern is operationally important because NUTMGS does not require experts to label every detected tree. A small expert-labeled seed set can train a strong street-view classifier, high-confidence predictions can expand labels over the GSV candidate pool, and experts can focus on rare or uncertain cases. Figure 5 tests the second half of that loop for remote sensing. The two panels show increasing label regimes: all 20 species at $k \in \{1, 10, 100\}$, and the five highest-support species at k=200–800, using the Sentinel-2 XGBoost model. Accuracy increases with the number of labels across species, and we do not observe accuracy plateauing for any species. This demonstrates the value of NUTMGS: bridging the gap in labels across tree species is also a route to bridging the gap in species-identification accuracy for tree species maps.

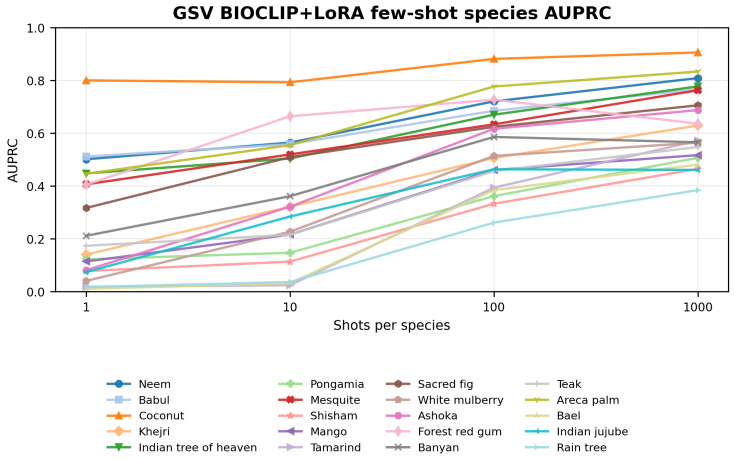


Figure 4: **Street-view few-shot species classification.** GSV BIOCLIP+LoRA curves are shown for the 20 species with binary species-vs-rest remote-sensing models.

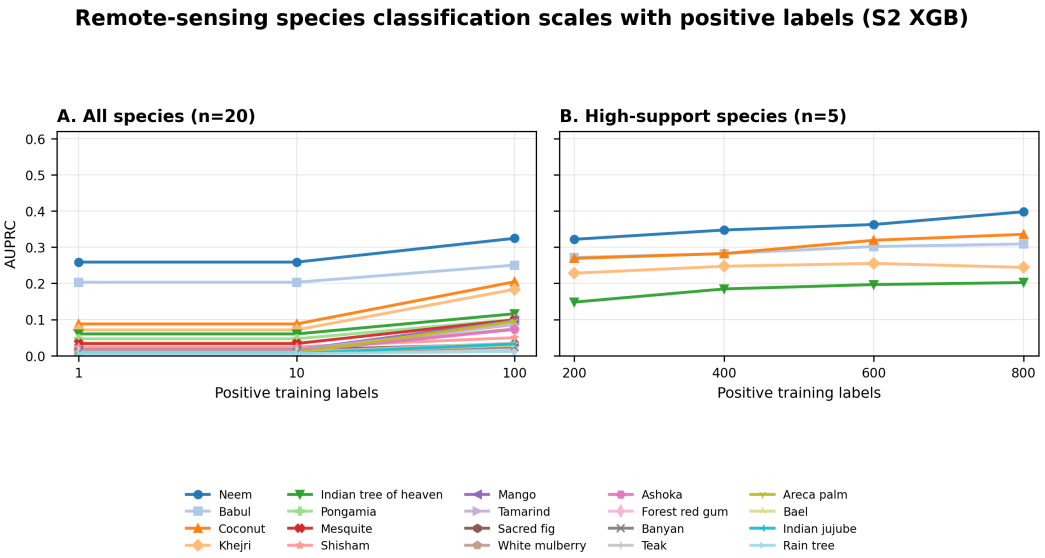


Figure 5: **Remote-sensing species classification improves with positive training labels.** Each panel uses the species with sufficient training positives for that label regime: all 20 species for $k \in \{1, 10, 100\}$ and five high-support species for k=200–800. Lines show individual species for the Sentinel-2 XGBoost model. Curves are not directly comparable across panels by absolute value because species sets differ, but within-panel slopes show scaling behavior within each support stratum.

4.4 Foundation models still require in-domain labels

Across tasks, foundation models improve some baselines but do not change the core ranking. For detection, Grounding DINO improves recall but has much lower state-balanced AP_{50} and precision than Faster R-CNN or Mask R-CNN (Table 2), while DINOv3 improves Plain-DETR but does not close the gap to the convolutional detectors. For image-based species classification, fine-tuned BIOCLIP improves over ResNet-50 on GSV, especially for rare species (Table 3), but zero-shot BIOCLIP-2 reaches only 0.262 AUPRC overall, compared to 0.335 for BIOCLIP+LoRA, and the top-5 gap is larger (0.471 vs. 0.809). This indicates that in-domain labels have a larger impact on accuracy than the difference between BIOCLIP-2 and a supervised ResNet baseline. SkySat foundation-model results are mixed and remain weak in absolute AUPRC. For geospatial embeddings, PRESTO, TESSERA, and AlphaEarth still do not beat the strongest supervised MLP or XGBoost

baselines in Table 3. The main conclusion is that Agroforestree is not solved by substituting larger pretrained models. Task- and modality-specific supervision still dominates performance.

5 Discussion

What Agroforestree enables. The main bottleneck for tree species mapping is not model architecture alone, but the absence of geometrically precise species labels at scale. Agroforestree addresses this bottleneck by showing that NUTMGS can turn sparse expert species knowledge into scalable labels for new tree species. The evaluation shows that the required components are tractable: individual-tree detection from SkySat works across Indian states (Table 2), and species classification is strongest with ground-level imagery but satellite-only species ID is tractable. Sentinel, PlanetScope, and combined Sentinel–PlanetScope time series outperform SkySat at top-5 despite coarser spatial resolution.

The key contribution is the individual-tree linkage across five modalities. SkySat provides the spatial resolution needed to detect individual crowns. GSV provides the visual detail needed for accurate species labels. PlanetScope provides dense temporal sampling, Sentinel-2 adds spectral information, and Sentinel-1 provides cloud-robust radar measurements. Together, these modalities make it possible to move from expert labels on individual trees to wall-to-wall species mapping.

The larger goal is to make understudied agroforestry species measurable enough for AI4Science. Species-resolved, longitudinal remote-sensing data across environments would enable agronomic studies of which tree and crop species farmers are actually using, climate studies of how different species respond to heat and water stress, and ecological studies of species populations that are otherwise invisible in public data. Agroforestree provides the first step: a dataset and labeling pipeline for building those species-level measurements in landscapes where field inventories, municipal tree censuses, and existing biodiversity platforms do not provide enough labels.

Limitations. NUTMGS is currently bottlenecked by SkySat and street-view coverage, both of which are controlled by private providers. We reduce these constraints by releasing authored 360° imagery and collection protocols for ground views, and by releasing a public VHR satellite tree detection dataset, but the dependency remains. Future work should reduce this dependence by making individual-tree detection work from PlanetScope-scale imagery with global coverage and by supporting researcher-collected 360° imagery instead of GSV.

Species labels also follow a power-law distribution, which makes it difficult to obtain enough positives for most species under natural sampling. The high AUPRC lifts in Appendix Table 7 suggest a practical way forward: use current models to supersample unlabeled candidate pools so new annotation batches are enriched for the target species.

Finally, current satellite species metrics are not yet sufficient for operational maps for many species. The path to operational mapping is to scale NUTMGS species by species, use GSV and expert labels to expand the training set, and continue adding labels until satellite accuracy plateaus at a map-ready level.

6 Ethics and Broader Impact

Fairness and representation. A central motivation for Agroforestree is that existing tree species benchmarks exist only where municipal tree inventories (Auto-Arborist) or forest management data (TreeSatAI) already do, excluding the rural smallholder regions where most agroforestry occurs. Producing individual-tree species labels from street-view and VHR satellite coverage broadens the set of regions where tree monitoring can be built.

Data and Code Availability

We will release the species-identification dataset at `https://source.coop/planet/agroforestry-tree-species-identification-india` and the individual-tree detection dataset at `https://source.coop/planet/`

agroforestry-individual-tree-detection-india. Code will be released at https://github.com/siddsach/SpeciesMapping.

References

Sophia Ahlswede et al. Treesatai benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing. *Earth System Science Data*, 15:681–706, 2023.

S.A. Al-Emadi, Y. Yang, and F. Ofli. Benchmarking object detectors under real-world distribution shifts in satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

AlphaEarth team. AlphaEarth: A global foundation model for earth observation. *arXiv preprint*, 2025. To appear.

Sara Beery, Elan Cole, Armand Gjoka, et al. The auto arborist dataset: A large-scale benchmark for multiview urban tree classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21173–21183, 2022.

Agnès Bégué, Damien Arvor, Beatriz Bellon, Julie Betbeder, Diego de Abelleyra, Rodrigo P.D. Ferraz, Valentine Lebourgeois, Camille Lelong, Margareth Simões, and Santiago R. Verón. Remote sensing and cropping practices: A review. *Remote Sensing*, 10(99), 2018.

Martin Brandt, Compton J. Tucker, Ankit Kariryaa, Kjeld Rasmussen, Christin Abel, Jennifer Small, Jerome Chave, Laura Vang Rasmussen, Pierre Hiernaux, Abdoul Aziz Diouf, et al. An unexpectedly large count of trees in the west african sahara and sahel. *Nature*, 587:78–82, 2020.

Rémi Cardinael, Georg Cadisch, Marie Gosme, Maren Oelbermann, and Meine van Noordwijk. Climate change mitigation and adaptation in agriculture: Why agroforestry should be part of the solution. *Agriculture, Ecosystems & Environment*, 319:107555, 2021.

Jérôme Chave, David Coomes, Steven Jansen, Simon L. Lewis, Nathan G. Swenson, and Amy E. Zanne. Towards a worldwide wood economics spectrum. *Ecology Letters*, 12(4):351–366, 2009.

Grace J. Di Cecco, Vijay Barve, Michael W. Belitz, Brian J. Stucky, Robert P. Guralnick, and Allen H. Hurlbert. Observing the observers: How participants contribute data to inaturalist and implications for biodiversity science. *BioScience*, 71(11):1179–1188, 2021.

Fabian Ewald Fassnacht, Hooman Latifi, Krzysztof Sterenczak, Aneta Modzelewska, Michael Lefsky, Lars T. Waser, Christoph Straub, and Aniruddha Ghosh. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186:64–87, 2016.

Google Research. Auto-Arborist dataset: project page. https://google.github.io/auto-arborist/, 2025. "Due to maintenance constraints, the dataset was turned down on May 27, 2025." Accessed April 2026.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

iNaturalist. Observe locally, identify globally? iNaturalist blog post on global coverage patterns. https://www.inaturalist.org/blog/27687, 2019. Documents iNaturalist's North American observation bias and discusses global coverage heterogeneity.

iNaturalist. iNaturalist: A community for naturalists. https://www.inaturalist.org, 2024. Citizen-science species observation platform.

Roeland Kindt, Lars Graudal, Jens-Peter B. Lillesø, Fabio Pedercini, Paul Smith, and Ramni Jamnadass. GlobalUsefulNativeTrees, a database documenting 14,014 tree species, supports synergies between biodiversity recovery and local livelihoods in landscape restoration. *Scientific Reports*, 13 (1):12640, 2023. doi: 10.1038/s41598-023-39552-1.

Jordi Laguarta Soler, Thomas Friedel, and Sherrie Wang. Combining deep learning and street view imagery to map smallholder crop types. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision (ECCV)*, 2014.

Yutong Lin, Yuhui Yuan, Zheng Zhang, Chen Li, Nanning Zheng, and Han Hu. DETR doesn't need multi-scale or locality design. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

P.K. Ramachandran Nair and Dennis Garrity, editors. *Agroforestry – The Future of Global Land Use*. Springer, Dordrecht, 2012.

Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5):523, 2019.

Planted team. Planted: A dataset for planted forest identification from multi-satellite time series. *arXiv preprint arXiv:2406.18554*, 2024.

PureForest team. PureForest: A large-scale aerial lidar and aerial imagery dataset for tree species classification in monospecific forests. *arXiv preprint arXiv:2404.12064*, 2024.

Gyri Reiersen, David Dao, Björn Lütjens, Konstantin Klemmer, Kenza Amara, Attila Steinegger, Ce Zhang, and Xiaoxiang Zhu. Reforestree: A dataset for estimating tropical forest carbon stock with deep learning and aerial imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, 2015.

Stephanie Roe, Charlotte Streck, Michael Obersteiner, et al. Contribution of the land sector to a 1.5°c world. *Nature Climate Change*, 9(11):817–828, 2019.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025.

Samuel Stevens, Jiaman Wu, Matthew J. Thompson, Elizabeth G. Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M. Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun Chao, and Yu Su. BIOCLIP: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

TESSERA team. TESSERA: A foundation model for sentinel-1 and sentinel-2 time series. *arXiv preprint*, 2025. To appear.

Sule Nur Topgul, Elif Sertel, Samet Aksoy, Cem Unsalan, and Johan E.S. Fransson. Vhrtrees: a new benchmark dataset for tree detection in satellite imagery and performance evaluation with yolo-based models. *Frontiers in Forests and Global Change*, 7, 2025.

Gabriel Tseng, Ivan Zvonkov, Catherine L. Nakalembe, and Hannah Kerner. CropHarvest: A global dataset for crop-type classification. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2021.

Gabriel Tseng, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023.

Josh Veitch-Michaelis, Andrew Cottam, Daniella Schweizer, Eben N. Broadbent, David Dao, Ce Zhang, Angelica Almeyda Zambrano, and Simeon Max. Oam-tcd: A globally diverse dataset of high-resolution tree cover maps. *arXiv preprint arXiv:2407.11743*, 2024.

Ben G. Weinstein, Sergio Marconi, Stephanie Bohlman, Alina Zare, and Ethan White. Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11:1309, 2019.

Xiangyu Zhao, Yicheng Chen, Xilin Li, Yuchen Liang, Jingbo Wang, Feng Wu, and Wenming Zhang. Grounding DINO: Marrying DINO with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*, 2023.

Robert J. Zomer, Henry Neufeldt, Jianchu Xu, Antje Ahrends, Deborah Bossio, Antonio Trabucco, Meine van Noordwijk, and Mingcheng Wang. Global tree cover and biomass carbon on agricultural land: The contribution of agroforestry to global and national carbon budgets. *Scientific Reports*, 6: 29987, 2016.

A Detection: additional results

Table 4: Per-state AP_{50} for the state-balanced detection benchmark. States with at least 10 validation images are included.

State	Images	Faster R-CNN	Mask R-CNN	Grounding DINO	Plain-DETR	DINOv3+DETR
Assam	37	46.6	47.1	29.2	31.2	35.1
Bihar	49	57.8	56.6	36.0	35.5	38.3
Chhattisgarh	47	71.5	71.4	36.9	40.3	43.2
Goa	53	58.8	59.3	32.0	34.3	36.4
Gujarat	39	67.1	65.0	39.0	39.5	40.9
Haryana	42	81.4	81.8	50.5	56.7	56.1
Jharkhand	49	65.2	64.5	38.4	40.3	44.4
Karnataka	1144	62.5	62.3	67.3	64.5	63.9
Madhya Pradesh	58	67.0	65.8	42.3	42.6	45.8
Maharashtra	48	65.7	62.9	40.1	41.2	43.7
Odisha	41	69.5	69.6	40.4	44.3	48.1
Punjab	51	60.6	62.3	34.5	34.6	37.8
Rajasthan	805	69.4	69.8	73.1	66.7	69.2
Tamil Nadu	42	61.9	63.1	38.1	38.1	41.2
Uttar Pradesh	105	63.0	63.3	36.6	37.8	41.5
Uttarakhand	50	65.8	65.3	32.2	33.9	35.6
West Bengal	33	50.6	48.4	30.2	29.2	35.3

Table 5: Per-state tree-count R^2 for the state-balanced detection benchmark. Count R^2 is squared Pearson correlation between per-image annotated and predicted tree counts.

State	Images	Faster R-CNN	Mask R-CNN	Grounding DINO	Plain-DETR	DINOv3+DETR
Assam	37	0.80	0.78	0.82	0.78	0.83
Bihar	49	0.83	0.84	0.92	0.79	0.82
Chhattisgarh	47	0.70	0.75	0.77	0.37	0.49
Goa	53	0.90	0.91	0.89	0.74	0.79
Gujarat	39	0.92	0.87	0.78	0.63	0.59
Haryana	42	0.82	0.86	0.68	0.63	0.67
Jharkhand	49	0.63	0.65	0.63	0.69	0.80
Karnataka	1144	0.54	0.55	0.60	0.54	0.56
Madhya Pradesh	58	0.78	0.77	0.74	0.70	0.79
Maharashtra	48	0.74	0.75	0.84	0.44	0.41
Odisha	41	0.78	0.79	0.81	0.74	0.63
Punjab	51	0.75	0.68	0.66	0.69	0.84
Rajasthan	805	0.77	0.77	0.76	0.83	0.81
Tamil Nadu	42	0.81	0.85	0.66	0.75	0.75
Uttar Pradesh	105	0.78	0.83	0.86	0.73	0.77
Uttarakhand	50	0.82	0.81	0.88	0.76	0.80
West Bengal	33	0.81	0.72	0.91	0.85	0.91

B Species: additional results

C Cross-view geolocalization: matching accuracy

Table 6: Per-species street-view one-vs-rest AUPRC on the held-out test set for the 20 species used in binary remote-sensing classification. Test positives is the number of positive examples for that class in the held-out test set. Per-species BIOCLIP-2 zero-shot values are omitted because zero-shot results are summarized only at aggregate level in Table 3.

Species	Test positives	ResNet-50	CLIP+LoRA	BIOCLIP+LoRA
Neem	460	0.76	0.82	0.84
Babul	361	0.72	0.80	0.82
Coconut	156	0.84	0.90	0.90
Khejri	127	0.54	0.64	0.69
Ailanthus	107	0.66	0.79	0.79
Pongamia	83	0.45	0.59	0.55
Mesquite	60	0.67	0.70	0.78
Shisham	43	0.26	0.43	0.45
Mango	29	0.32	0.50	0.52
Tamarind	29	0.25	0.47	0.56
Sacred fig	28	0.45	0.51	0.67
White mulberry	27	0.55	0.52	0.65
Ashoka	21	0.45	0.65	0.72
Forest red gum	18	0.48	0.67	0.76
Banyan	18	0.33	0.42	0.59
Teak	17	0.24	0.42	0.49
Areca palm	14	0.53	0.73	0.77
Bael	13	0.16	0.36	0.42
Indian jujube	11	0.15	0.48	0.50
Rain tree	10	0.14	0.31	0.38

Table 7: Per-species AUPRC for a fixed street-view model and a fixed remote-sensing model on the 20 species with binary species-vs-rest remote-sensing runs. n and prevalence are measured on the held-out one-vs-rest test set. RS lift is S1+S2+PlanetScope XGBoost AUPRC divided by prevalence.

Species	n	Prev. (%)	GSV BIOCLIP+LoRA	S1+S2+PS XGB	RS lift
Azadirachta indica	460	25.9	0.84	0.58	2.2
Vachellia nilotica	361	20.3	0.82	0.49	2.4
Cocos nucifera	156	8.8	0.90	0.54	6.2
Prosopis cineraria	127	7.1	0.69	0.33	4.7
Ailanthus excelsa	107	6.0	0.79	0.32	5.4
Pongamia pinnata	83	4.7	0.55	0.28	6.0
Prosopis juliflora	60	3.4	0.78	0.23	6.8
Dalbergia sissoo	43	2.4	0.45	0.10	4.2
Tamarindus indica	29	1.6	0.56	0.33	20.4
Mangifera indica	29	1.6	0.52	0.13	8.0
Ficus religiosa	28	1.6	0.67	0.10	6.4
Morus alba	27	1.5	0.65	0.14	9.1
Saraca asoca	21	1.2	0.72	0.13	10.8
Ficus benghalensis	18	1.0	0.59	0.13	13.0
Eucalyptus tereticornis	18	1.0	0.76	0.03	3.2
Tectona grandis	17	1.0	0.49	0.07	7.1
Areca catechu	14	0.8	0.77	0.46	57.9
Aegle marmelos	13	0.7	0.42	0.04	4.9
Ziziphus jujuba	11	0.6	0.50	0.11	18.3
Samanea saman	10	0.6	0.38	0.02	3.5

Table 8: Cross-view geolocalization matching accuracy on a 7,380-pair human-validated subset.

Category	Correct	Incorrect	Accuracy	Total
All	6167	1213	0.836	7380
Near	4580	851	0.843	5431
Far	1587	362	0.814	1949