

PlantationBench: a multiscale, multimodal remote sensing benchmark for plantation mapping under distribution shift

Angela Tsao

Department of Earth System Science
Stanford University, Stanford, CA, 94305

angtsao@stanford.edu

David Lobell

Department of Earth System Science
Stanford University, Stanford, CA, 94305

dlobell@stanford.edu

Abstract

Satellite-based land cover maps fail to distinguish between natural forests and planted trees when certifying agricultural plantations as deforestation-free, a critical need for smallholder farmers facing emerging legislation around sustainable supply chains (such as the European Union Deforestation Regulation). Previous efforts to map tree plantations exclude smallholder farms, which require intra-field detail about individual tree crowns that can only be observed in costly, very-high resolution (VHR, $\leq 1m$) satellite imagery, and limit scalability. To provide a replicable benchmark for mapping tree crop plantations, we introduce a novel earth observations dataset of hierarchically-labeled plantation and forest locations across the African continent. We compile for these points a unique feature set combining a VHR satellite-derived canopy height product and open-source multispectral and synthetic aperture radar data, to fuse high spatial resolution with multitemporal, multiscale, and multimodal features. Our dataset has continental coverage for Africa’s plantation-growing nations across a diverse range of climate zones, with more delineation for species and geographical region than any other related tree dataset. Samples are structured with different domains to account for real-world distribution shifts to study the interplay of domain generalization with various feature attributes of remotely sensed data. We test how pre-trained models can perform on the task of tree plantation mapping, and benchmark performance at scale and across different distributions.

1. Introduction

The 2023 European Union Regulation on Deforestation-Free Products (EUDR) imposes sustainability mandates on \$110 billion of global trade across 7 of the world’s key commodities, with several other agricultural products under consideration for later updates to the policy [17]. EUDR

imposes one key constraint for imported commodities: they must be certified “deforestation-free”. Following a few key definitions, deforestation-free means that

the relevant products contain, have been fed with or have been made using, relevant commodities that were produced on land that has not been subject to deforestation after 31 December, 2020,

where deforestation includes conversion of forest land to “agricultural plantations.” Here, we use the term tree plantations interchangeably with agricultural plantations along the same EU criteria: where “agricultural plantation” means

land with tree stands in agricultural production systems, such as fruit tree plantations, oil palm plantations, olive orchards and agroforestry systems where crops are grown under tree cover; it includes all plantations of relevant commodities other than wood [17].

Tree plantations, colloquially defined as areas in which trees were planted by humans, are differentiated from forests by economic function and the central role of human management. They are a vital tool for sustainable development [25, 34, 35, 41, 51] due to the economic value they bring to smallholder (<2ha) farmers, who comprise as much as 70% of Africa’s population [9]. Tree plantations are also a major driver of deforestation, accounting for as much as 37% of forest loss in national-scale protected areas [26] and 7% of all deforestation in Africa [31]. However, there is a critical impediment to understanding the complex interplay between tree plantations and deforestation: existing global land cover and land use maps fail to differentiate between various types of tree cover. As a consequence, tree plantations existing since 2020 (the base year) may nonetheless be flagged as a deforested plot if misclassified in the base year, or if change detection methods flag management practices such as thinning or re-planting. Current compliance efforts rely on existing maps like Hansen

et al. [23]’s Global Forest Change map, which have been criticized for their failure to capture plantations as a distinct form of land use and land cover [49], as well as inconsistencies for small patches <2ha [33]. While more institutional farmers are able to hire consultants to help with certification, smallholder farmers do not have the resources to register their farms and their continued inclusion in the global markets may depend on the availability of open-source certification tools. These smallholders are most vulnerable to the administrative costs of compliance [8]. Equitable, scalable enforcement of emerging policies like the EUDR requires globally consistent, accurate mapping of plantation cover in the base year. We introduce PlantationBench to support these urgent smallholder plantation mapping efforts; most recently, officials postponed enforcement of the EUDR due to the complexity of plantation certification.

While the task of plantation mapping is deeply valuable in sustainability and environmental policy applications, it is also an interesting challenge for the computer vision domain. In particular, the question of domain generalization has motivated other real-world image datasets [6, 28], and it is especially important for understanding the complexity of generalization when using multiple remote sensing modalities. Remote sensing data deployed at large spatial scale are vulnerable to changes across space, due to many real-world shifts such as climate, aridity, and land use, all of which make training continental or global models difficult [11]. In addition, learning multiscale features within and across multimodal sources of data is a prominent challenge to benchmark for emerging models [20, 42]. We present PlantationBench as a grounded benchmark task that can be used to evaluate how well existing foundation models can perform on real tasks involving complex multimodal, multiscale inputs that differ in feature processing compared to the conventional training sets for earth observation foundation models. Subclass labels for tree plantation crop type unlock additional value as a relevant challenge in hierarchical multi-label classification. This task has been increasing in interest [30, 55, 60], with excitement about methods ranging from visual understanding and consistency [2] to CLIP-based hierarchical text-based (label) metadata [43].

In summary, with PlantationBench, we make these contributions:

- We develop pipeline scripts for combining known tree plantation locations with multiple streams of remote sensing data, to generate a lightweight yet large-scale dataset with curated features for agricultural plantation identification. This processing pipeline enables inference and classification of farmers’ land based on only geo-coordinates as input.
- We introduce the PlantationBench dataset, upon which models can be trained and evaluated, to contribute directly to certifying smallholder tree plantations as

deforestation-free for policies like the EUDR.

- We demonstrate that diverse coverage of crop type and geography benefit model generalization, and that both multimodal and multiscale data bring complementary improvement.

2. Related Work

2.1. Land cover mapping

Concerted efforts have been made to map natural forest land cover, because forests are recognized for their high environmental value. Most existing land cover maps include a “forest” class, typically based on the biophysical definition that forests meet $\geq 10\%$ tree cover threshold on at least .5 ha contiguously. This parallels the FAO definition, but does not capture the land use element of agricultural plantations. For example, Shimada et al. [46] use synthetic aperture radar (SAR) to generate an explicit forest/non-forest map (PALSAR FNF), but note that various types of plantation cover are included in the forest category, as computed metrics based on the HV-polarization band from SAR fail to distinguish between plantation and forest. The focus of PALSAR FNF is not on trees, however, so this is reasonable. Similarly, the well-established Global Land Cover and Land Use Change dataset, which is a leading reference for mapping both tree cover and agricultural use, does not have separate provisions to categorize tree plantations as agricultural land cover [40]. Meanwhile, the Hansen Global Forest Change (GFC) dataset, which establishes forest loss and deforestation at global scale using Landsat data, also sticks strictly to defining forest area by tree cover density, canopy height, and contiguous area [23]. Even patches which do not fit the biophysical criteria outlined in Hansen et al. [23], such as pineapple plantations shorter than the minimum height threshold, get classified as forest by GFC [49]. In sum, none of the existing global, multi-crop products are capable of mapping plantation separately from forest. Tab. 1 identifies the key global tree-mapping products and their provisions to address plantation cover as a distinct class from forest.

Emerging satellite-based mapping of specific crops, such as oil palm, for individual nations or regions have achieved success [15], but there are major discrepancies across regions. Efforts to map at geographical scale introduce complex distribution shifts in data, where for new locations, model performance degrades and transfer learning is difficult [6, 28, 59]. One large-scale dataset addressing the challenge of mapping at geographical scale focuses on classifying timber plantation forests [39], which is a distinct but complementary task to mapping agricultural tree plantations, which produce a non-timber crop. Pazos-Outón et al. [39] point out the need for multiscale, multimodal, and multitemporal data sources for global or continental mapping.

Besides satellite-based mapping approaches, large-scale

surveys have also been undertaken to account for trees, typically at a national scale. Within national forest inventories, inconsistent definitions, data quality, and infrequent updates may also result in inaccurate maps. For example, official reports of cocoa plantation cover in Ghana were shown to underestimate areas with planted trees by 40% [26].

2.2. Agricultural tree plantation mapping

The first global-scale attempt to map tree plantations separately from natural forest was compiled in 2015 as the Spatial Database of Planted Trees (SDPT). It was estimated that tree crops cover at least 50 million ha, while timber plantations span 170 million ha, 4% of global forest cover [57]. However, the SDPT is incomplete in coverage. It only maps a subset of countries which are known to have high levels of plantation forest, and its combination of methods (manual polygon delineation, random forest-based supervised classification, and mixed methods) are cumbersome, difficult to scale, and somewhat disaggregated based on national-level datasets. It is difficult and time-consuming to update such a reference map. SDPTv2.0, an updated version of SDPT was released in 2024 attempting to improve the coverage and recency of source data [44]. It includes several more countries in sub-Saharan Africa. However, the data sources they aggregate remain outdated, even for countries that are newly added, with many sources dating back as far as 2013. Due to resource and method limitations, SDPTv2.0 still fails to achieve a spatially and temporally consistent map of tree plantations. This renders the map far less useful for applications such as tree crop traceability, which requires knowledge of plantation status at a specific time frame.

The United Nations' Food and Agriculture Organization collects country-level data about tree crop production, but these may face severe under-reporting, and production statistics are too variable to estimate planted area from harvest quantity. Interest in specific crops has led to the creation of single-crop maps, such as for oil palm [15], coconut palm [16], cocoa [26], and cashew [58]. These crop-specific maps have leveraged a variety of linear and deep learning methods to predict plantation cover from both tabular and image satellite features, with different strategies for summarizing multitemporal and multispectral features. However, agricultural tree plantations are highly spectrally diverse across different species types, geographies, and plantation conditions, which makes a model trained for any given geography or given species difficult to transfer. For the single-crop or single-country mapping tasks, accuracies range from 75%-90% [15, 16, 26, 58]. This shows that plantation mapping is a difficult task even without considering domain generalization challenges for the multispecies, continental mapping task.

Fagan et al. [18] attempt to build a more general model to address the challenge of classifying plots as either natu-

ral forest or plantation for tropical regions, using medium-resolution Landsat, SAR, and Sentinel-1 features. However, their data in Africa is limited exclusively to palm species or "unknown", and exclude smallholder fields <.45 ha. Notably, user's and producer's accuracy on plantations in Africa are worse than in any other tested region.

2.3. Remote sensing features for trees

Most commonly, the application of computer vision and remote sensing data for individual trees has been toward the task of individual tree detection [7, 11, 61] or tree species classification. These tasks typically involve aerial [1, 3, 4] or street-view [6] imagery at more local scales. Critically, including multimodal inputs in modeling has been found to improve overall performance and generalization [6, 27, 39]. There has been a rise in the popularity of earth observation (EO) foundation models and associated benchmarking tasks [5, 14, 20, 29, 36, 42]. Notably, emerging datasets and foundation models focus on the challenge of learning multiscale features [20, 42, 47] within and across information-rich, multimodal sources of data [20–22, 29, 36]. Most recently, there has also been a focus on publishing location-based embeddings, fixed-length learned representation vectors extracted from a EO foundation model's various inputs [53, 54]. These embeddings provide a simple platform for supervised learning or unsupervised clustering.

At the same time, many studies have highlighted the complexity of tree species distribution, and the difficulties associated with the diversity of phenologies and appearances [56], which makes plantation classification and tree-related studies particularly salient as an EO evaluation task. For PlantationBench, we acquire high-quality plantation and forest locations based on field studies, which is an area that has previously been filled with noisy labels and inaccurate mapping efforts. Additionally, we contribute crop type labels for known tree plantations, realizing that sub-label shift is an important type of distribution shift, where tailedness and frequency of certain sub-classes (types of plantations) can make model performance highly variable across different regions.

3. PlantationBench Dataset

We publish the most policy-comprehensive multi-feature dataset for agricultural tree plantations, by emphasizing smallholder inclusion and crop type labels. To acquire plantation labels for training and testing, we aggregate existing sources of plantation geolocation data from SDPT field records [44], species-level maps [15, 16], and large-scale research studies [18]. Additionally, we collect new plantation and forest locations from field sites, manual labeling, and ground verification for each of the top-10 highest-production African countries and countries where our partner organizations have previous engagements. De-

| Land Cover Mapping | | | | | | | |
|---|---------------------|----------------------------|-----------|-------------------|--|--|--|
| Name | Extent | Plantation Provi- sions | Year | Resolution (m) | Methods | | |
| GLAD Global Land Cover and Land Use Change [40] | Global | Treated as forest | 2020 | 30 | Decision tree | | |
| ALOS Palsar Forest-not-forest [46] | Global | Treated as forest | 2007-2020 | 25 | Rule-based thresholding | | |
| Hansen Global Forest Change [23] | Global | Treated as forest | 2013-2023 | 30 | Decision trees (w bagging) | | |
| Spatial Database of Planted Trees [44] | Global (incomplete) | Separately labelled | 2015 | Vector | Mixed (sup. classification, manual, surveys) | | |

Table 1. AI-based tree cover datasets over Africa lack distinction for type of tree cover

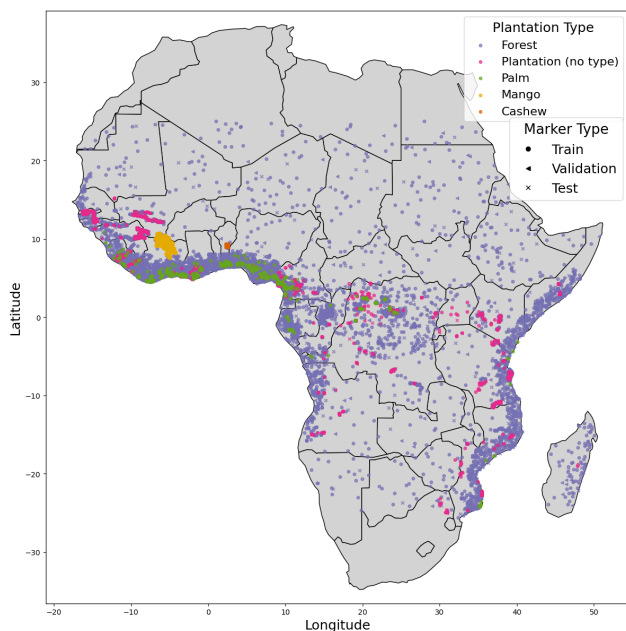


Figure 1. Label distribution across Africa.

tails about the collection process are described in Sec. 3.1.1 and aggregate summarization of the dataset is provided in Tab. 2 and Fig. 1. All points from “PlantationBench Manual”, Cashew, and Mango subsets are plantation locations collected by our team and partners. A minority (5.7%, 888/15347) of other plantation locations, specifically the Oil Palm [15], Coconut Palm [15], and Fagan [18] subsets, are previously published by researchers (Tab. 2).

For the African continent, even the best existing plantation mapping efforts face severe data limitations over Africa; Fagan et al. [18] includes 10+ species and crop types for other continents, but in Africa, exclusively sources species labels for oil palm and coconut palm for their training data. Any other species are referred to with a homo-

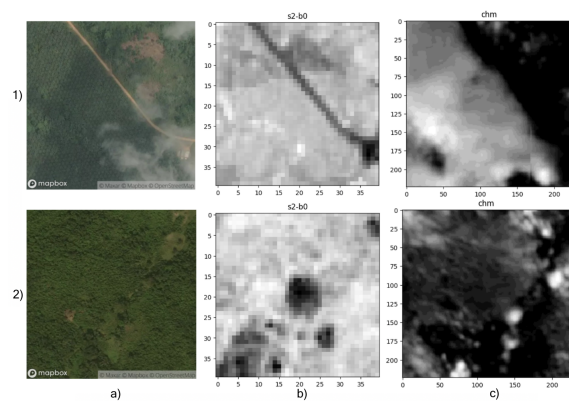


Figure 2. Processed feature input examples for 1) plantation (oil palm, West Africa) and 2) forest (West Africa) point examples. a) Very-high resolution aerial image for visual reference only, b) Sentinel-2 NDVI (400x400m x 4 channel), and c) VHR CHM (224mx224m x 1 channel).

geneous “mixed” class. Through PlantationBench, we significantly augment the geographical and crop type scope compared to existing works. In particular, while species labels are not comprehensively included in PlantationBench, we represent all major categories of commercial and subsistence tree crop, among fruit, nut, and oil product. This results in a two-tier tree hierarchy for multi-label classification, where tier 1 first solves the plantation-forest mapping and tier 2 addresses crop type classification.

3.1. Data

3.1.1. Annotation Collection and Geolocation-based Pipeline

The framework for processing samples involves collecting point coordinates from the interior of different plots. In the case of partner label data, these coordinates exist from field surveys where enumerators would visit farm sites, tagging class labels and recording GPS within tree plantations and

various natural forest locations. A subset of the dataset, called "PlantationBench Manual", was sourced by manually recording coordinates for tree plantations in known agriculturally-producing countries.

Besides publishing the image features for labeled samples in the PlantationBench dataset, we make openly available the scripts to process additional inference or training-ready features for any given list of input coordinates. This makes it possible for practitioners and researchers to input locations in need of classification and directly extract features ready for inference. The data processing pipeline deploys through Google Earth Engine's Python API, handling feature engineering, aggregation, and image exports for the trimodal feature set. The combined processed image features for example points are visualized in Fig. 2. We describe further details behind each of the input features.

3.1.2. Canopy Height Map

Canopy height maps at high spatial resolution can provide information about planting patterns and spatial structures associated with managed tree plantations. This includes row planting or cluster growth. However, to provide meaningful information at the smallholder level, it is valuable to have submeter imagery to delineate spacing between trees, which in structured plantations may be less than 10m between trunks. Tolan et al. [48] use Maxar Vivid2 VHR imagery and coupled dense aerial light detection and ranging (LiDAR) to predict canopy height over VHR images. These predictions are formed from training images primarily in 2020, although some images date back to 2017. The time frame allows applicability as a feature for the EUDR benchmarking task, as it assesses canopy height patterns circa 2020, from which plantations can be detected. It is also viable that future updated iterations of the global submeter CHM can be used for consistent monitoring of plantation cover change. For plantation and forest coordinates, we sample a 224m x 224m (224 x 224 pixel) patch from the CHM product, centered on the point latitude and longitude, where pixel value represents canopy height. This feature can be visualized in Fig. 2c and is treated as a grayscale image. Given the high cost and closed-source availability of very high resolution (VHR, <1m ground sampling distance) satellite data outside of US and Europe, we identify that secondary spatial products are a more accessible feature set at very high spatial resolution. The open-source availability of such emerging datasets encourages future use of aggregate product as an input feature for downstream tasks, but with the limitation that multispectral and multitemporal information must be supplemented with other data sources. One advantage of aggregating most-valuable features along these axes is that the filtered dataset is more computationally efficient for inference at scale.

3.1.3. Sentinel-2

Sentinel-2 Multispectral Instrument satellite data has multispectral information ranging from 10m resolution to 60m. The Sentinel-2 base data used in this study were sourced from the Level-2A orthorectified atmospherically corrected surface reflectance collection, acquired through the Google Earth Engine Data Catalog. We opt to use standard red and near infrared bands to compute a normalized difference vegetation index (NDVI) pixelwise for image patches [45]. In order to include multitemporal information which conveys seasonality, we composite all data from 2023 in 3-month windows to define a 4-season NDVI image time series. These windows are based on climatic dry and wet seasons, with quarterly cycles to accommodate intermediate seasons and East Africa's phenomenon of having two distinct wet seasons [38]. Phenological features are important because distinct crop types exhibit unique seasonal patterns. Each season is treated as a single channel in the stacked 4-channel feature. Based on the reference coordinates, we sample a 4-band, 400m x 400m (40 x 40 pixel) image centered on the point latitude and longitude, visualized in Fig. 2b.

3.1.4. Sentinel-1

Synthetic aperture radar (SAR) from the Sentinel-1 constellation provides an active sensing modality. SAR ground range detected scenes have the advantage of penetrating any visually obscuring features such as cloud cover, which presents significant difficulties for monitoring tropical regions like West Africa with exclusively optical satellite data [15, 31]. Following the example of other land cover research [18], we first sample 400m x 400m (40 x 40 pixel) patches centered at each point in our dataset. Then we process full-year composites of Sentinel-1 as tabular features at patch-level, by separately normalizing vertical-horizontal and vertical-vertical polarizations with computed entropy and aggregation statistics.

3.1.5. Foundation Model Embeddings

While extraction of satellite imagery is a critical step of many applications in computer vision for remote sensing, foundation model embeddings can be used as features for downstream tasks like classification. Embeddings are a foundation model's learned vector representations for a given location; they implicitly capture information from features used in training. Embeddings streamline steps such as feature engineering when used as input features to downstream supervised learning models. We export embeddings for PlantationBench sites from two recently published models. Google Satellite Embeddings provides a 64-band dataset at 10m spatial resolution, learning a single representation by leveraging leverages a deep Space Time Precision encoder over temporally sparse multimodal input sources including optical, radar, and LiDAR [12, 19]. On the other

hand, EarthIndex Embeddings generate 384-band vectors with 17.8m resolution (320 square meter patches) by passing the SoftCon [54] model over Sentinel-2 imagery.

4. Experiments

We conduct experiments to show the value added by PlantationBench’s specific contributions as a continental-scale, multimodal, multispecies dataset. Training procedure is standardized across experiments, with an allowance for 50 training epochs and default hyperparameter settings, described in more detail in supplemental. For all experiments, overall accuracy, weighted F1 score, precision, and recall were computed. Supplementary to experimenting with the custom PlantationBench images, we use our geo-referenced labels to extract embeddings for the downstream classification task, and directly train a random forest algorithm on embedding vectors from emerging EO foundation models.

4.1. Multimodality

We begin experiments to show that multiple feature streams in PlantationBench are valuable toward the plantation classification task. We default to a continental-scale train-test split and fine-tune a simple convolutional neural network on each of the individual feature sets. Then, we consider the multiscale feature set (VHR canopy height map and coarser multispectral Sentinel-2 images) and the full multiscale, multimodal feature set which adds radar features. We incorporate these different streams of data using a late fusion approach, where individual feature extractors are used on each input stream and their latent representations are then passed through a fully-connected layer to fuse outputs for the classification task [52]. We report that the multiscale synthesis of RGB features with NDVI time series increases classification accuracy significantly, and further addition of a different modality of data (synthetic aperture radar (SAR) from Sentinel-1) has some additional marginal benefit. The value of SAR backscatter reported in other plantation mapping projects may be most useful for classifying various palm cultivars. The different features offer distinct advantages depending on plantation type.

4.2. Baseline Results

Given the task of binary classification from multiple streams of remote sensing data, we adapt existing convolutional backbones to individually extract features for each input imagery. The resulting deep features are concatenated and input to a linear classification head which outputs class probabilities. We test how pre-trained backbone performs and find that initializing with pre-trained weights (based on ImageNet) is effective at improving accuracy by 3.5% compared to random weights (Tab. 4).

We test the value of transformer-based methods over traditional convolutional approaches for the plantation map-

ping task. In particular, given the success of pre-training for ResNet, we fine-tune Sat-MAE on the task but find that it slightly underperforms the convolutional models (Tab. 4). We note that the treatment of canopy height and vegetation index images as separate inputs may add less value compared to a multimodal attention mechanism for models pre-trained on coupled imagery. The embeddings-based baseline show promise for deploying foundation models on downstream tasks, as Google embeddings outperform even the best custom architectures we test by 10% (Tab. 4). There is significant value from global-scale, multimodal source inputs. However, EarthIndex embeddings underperform all other architectures. We observe that the F-1 scores for the embeddings-based models are very comparable to the accuracy, suggesting the large-scale models have balanced performance across classes. This is likely a benefit from global-scale, general-purpose pre-training.

4.3. Distribution Shifts

To effectively map plantations at the continental scale for Africa involves geographical distribution shift (due to changing climatic zones, aridity gradients, and similar physical conditions) as well as accounting for different types and frequency distributions of species. In order to test robustness of our model to these changes, we run an experiment leaving out each component subset from the full PlantationBench dataset. For country-level “leave-one-out” experiments, the model is trained with training datasets from all countries besides the designated left-out country, and test performance is evaluated on that country’s test set (Fig. 3).

We consider countries with sufficient data for fine-tuning (>500 samples total) and contrast the test performance from training on that country’s training set, all countries from the same region, and finally, all available data. Consistently, the model trained on all available data outperforms the more constrained models, and averages .80-.97 balanced F1-score across countries with >500 data points (Fig. 4).

4.3.1. Geographical distribution shift

For major tree crop-producing countries, we demonstrate a generalization gap between self-test (training on dataset including the country, testing on held-out set from the country) and transfer performance of a model trained in-country and tested on all others. We note that some countries generalize fairly well to others, but overall, the full-training case is the only case that generalizes over all countries.

4.3.2. Species distribution shift

We experiment with partitioning the data by crop type category, and assess how well a model trained without same-crop information does on the various tasks. We find that having same-crop examples is critical for classification accuracy, but some of the difference is mitigated with greater geographical representation. The crop types for which we

| Dataset | Region | # Plantations | # Forests | Plantation Proportion | Species |
|----------------------------------|------------------------|---------------|-----------|-----------------------|---------|
| PlantationBench Manual | East Africa | 528 | 1021 | .34 | Multi |
| PlantationBench Manual | West Africa | 1582 | 2440 | .39 | Multi |
| PlantationBench Manual | Central Africa | 458 | 1001 | .31 | Multi |
| Cashew | Benin | 7283 | 7305 | .50 | Nut |
| Mango | Cote d'Ivoire | 4608 | 0 | 1.0 | Fruit |
| Coconut Palm [16] | Continental | 69 | 1572 | .04 | Palm |
| Oil Palm [15] | Continental | 685 | 2579 | .21 | Palm |
| Fagan [18] | Continental (Tropical) | 134 | 1380 | .09 | Palm |
| PlantationBench Aggregate | Continental | 15347 | 17298 | .47 | Multi |

Table 2. PlantationBench data inclusions by source, crop type, and region

| Test Subset | CHM | Sentinel-2 | Sentinel-1 | CHM + S-2 (multiscale optical) | CHM + S-2 + S-1 (multiscale, multimodal) |
|----------------------------------|-----|------------|------------|-----------------------------------|---|
| PlantationBench Manual | .87 | .75 | .70 | .89 | .90 |
| Cashew | .73 | .82 | .49 | .85 | .85 |
| Coconut Palm | .96 | .94 | .83 | .96 | .96 |
| Oil Palm | .87 | .84 | .82 | .90 | .90 |
| Tropical Palm | .95 | .91 | .86 | .95 | .96 |
| PlantationBench Aggregate | .81 | .79 | .74 | .83 | .83 |

Table 3. Plantation classification results (overall accuracy) for different feature inclusions

| Method | Backbone | Acc | F1 |
|---------------|------------------------|-----|-----|
| ResNet Tri | ResNet-18 (Pretrained) | .84 | .80 |
| ResNet Tri | ResNet-18 (Random) | .81 | .80 |
| SatMAE Bi | ViT (Pretrained) | .81 | .72 |
| Google FM | AlphaEarth [12] | .95 | .94 |
| EarthIndex FM | Softcon [54] | .75 | .75 |

Table 4. Results from full training dataset with pre-training and different architectures (Tri (trimodal) includes canopy height, Sentinel-1, and Sentinel-2 features while Bi (bimodal) includes only image-based canopy height and Sentinel-1. FM (Foundation model embeddings) are based on the origin model’s training process, so Google includes 3+ modalities while EarthIndex relies on optical imagery.

had restricted geographies (cashew, mango in single country) experienced the most dramatic drop in performance (Fig. 5).

5. Discussion

5.1. Inference

We developed a framework for which to assess the deforestation-free status of an input location based on its class in 2020. In order to illustrate the value of this in real-world enforcement, we consider the location of several known smallholder oil palm locations from [50]. We show the model and dataset’s value in verifying the deforestation-

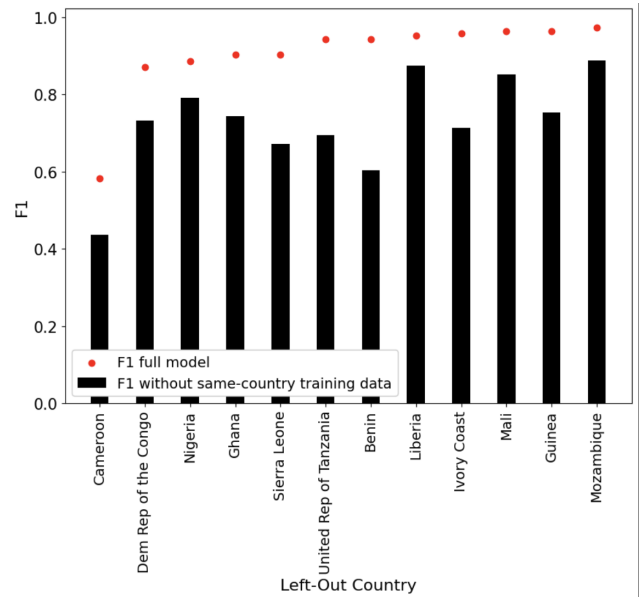


Figure 3. Model performance naturally declines when trained without same-country data. Heterogeneity of species composition is a driver, with countries like Benin and Ivory Coast, where select crops predominate the dataset, experiencing the greatest drop.

free compliance for smallholder farmers where traditional land cover maps from 2020 [46] misclassify agricultural tree plantations. Notably, these misclassifications are not

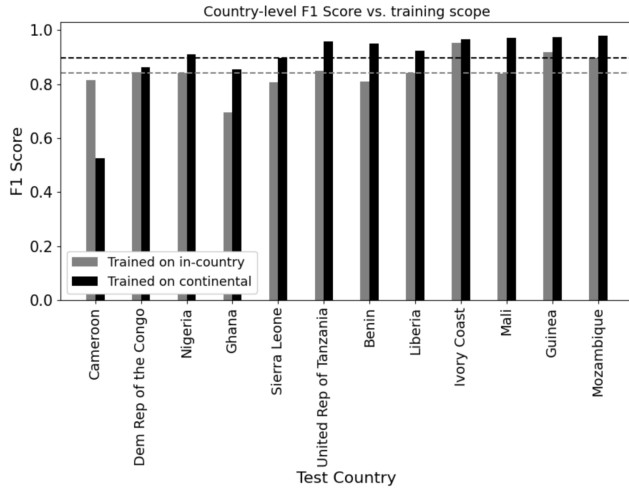


Figure 4. Model performance on individual country test sets improves when given access to data from other countries, with increased diversity of points in training data. Average test performance across all countries, for in-country and continental-scope training is represented with horizontal lines.

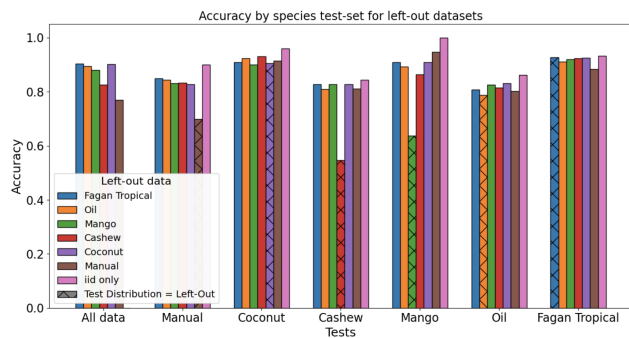


Figure 5. The removal of training data drawn from the same-distribution (same species) as a test subset (hatched bars) has the strongest negative effect on test accuracy on that subset compared to removing data from other sources.

pixel-level inconsistencies, but rather, errors over whole patches due to lack of a clear definition for plantation as non-forest tree cover. We consider a misclassification if over 50% pixels for a 10 x 10 pixel patch (25m/pixel) are classified as forest. Fig. 6 illustrates 3 examples where plantation locations given in [50] are classified as forest by the existing 2020-baseline forest-not-forest map, but are appropriately classified by our model.

5.2. Limitations

We present a simple modeling approach with baseline architectures and strategies for learning the plantation mapping task. Future work can improve on the strategy used to fuse the various input features and maximize joint learning over their unique traits of high-quality spatial, temporal,

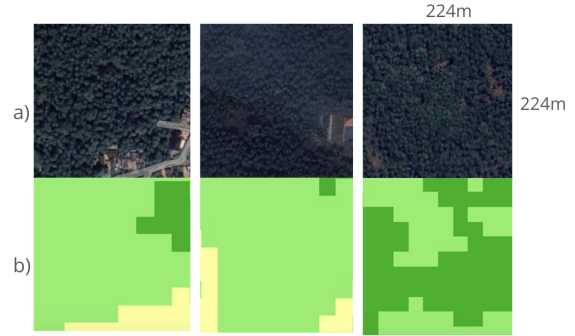


Figure 6. a) Very-high resolution aerial image for visual reference of example tree plantations classified correctly by our models, b) 2020 PALSAR Forest/Non-Forest land cover map [46] for same location. Dark green denotes "dense forest", light green is "non-dense forest", and yellow is "non-forest"

and spectral information. We show that supervised learning on foundation model general-purpose embeddings can work well and is an emerging practice with the potential to scale many earth applications. We hope to acquire more different subtypes (specific species) of fruit, nut, and oil crop among the crop types that we separated, as this will provide even more use cases for local governments, farmers, and exporters who are involved with smallholder supply chains. PlantationBench has serious implications for the future of equitable policy in sustainable development. EUDR and similar policies are necessary to protect global forests from irreversible damage, but it is imperative that technological solutions help empower resource-limited smallholder farmers who may otherwise be underprepared to comply with top-down policy.

5.3. Conclusion

Researchers have raced to develop foundation models for earth observation and understand how well these models transfer to downstream tasks. We show in our experiments that incorporating multimodal inputs allows for better learning of the end task, and that global-scale multimodal FM embeddings do well on the task. The first wave of remote sensing foundation models largely handled single-sensor inputs [24] with limited flexibility for multiscale inputs [37]. The next step with PlantationBench is to encourage and benchmark performance for methods that natively handle multimodal and multiscale imagery, particularly emerging EO foundation models such as SatMAE++ [37], GFM [32], and FoMoNet [10], among many other large-scale EO studies [4, 5, 13, 14, 20, 21, 29, 36, 42, 60]. PlantationBench provides an economically grounded, socially relevant, and timely application through which innovative research can be applied to solve a real-world challenge that affects global climate efforts and the livelihoods of millions of smallholder farmers.

References

- [1] Daniel Amigo, David Sánchez Pedroche, Jesús García, and José M Molina. Automatic individual tree detection from combination of aerial imagery, lidar and environment context. In *16th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2021)*, pages 294–303. Springer, 2022. 3
- [2] Anonymous. Visually consistent hierarchical image classification. In *The Thirteenth International Conference on Learning Representations*, 2025. 2
- [3] Méline Aubry-Kientz, Anthony Laybros, Ben Weinstein, James GC Ball, Toby Jackson, David Coomes, and Grégoire Vincent. Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:3927–3936, 2021. 3
- [4] Bulut Aygunes, Ramazan Gokberk Cinbis, and Selim Aksoy. Weakly supervised instance attention for multisource fine-grained object recognition with an application to tree species classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176:262–274, 2021. 3, 8
- [5] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. Atlaspretrain: A large-scale dataset for remote sensing image understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16772–16782, 2023. 3, 8
- [6] Sara Beery, Guanhang Wu, Trevor Edwards, Filip Pavetic, Bo Majewski, Shreyasee Mukherjee, Stanley Chan, John Morgan, Vivek Rathod, and Jonathan Huang. The auto arborist dataset: A large-scale benchmark for multiview urban forest monitoring under domain shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21294–21307, 2022. 2, 3
- [7] Mirela Beloiu, Lucca Heinzmann, Nataliia Rehus, Arthur Gessler, and Verena C Griess. Individual tree-crown detection and species identification in heterogeneous forests using aerial rgb imagery and deep learning. *Remote Sensing*, 15(5):1463, 2023. 3
- [8] Laila Berning and Metodi Sotirov. The coalitional politics of the european union regulation on deforestation-free products. *Forest Policy and Economics*, 158:103102, 2024. 2
- [9] M Biteye. 70% of africans make a living through agriculture, and technology could transform their world. In *World Economic Forum on Africa*, 2016. 1
- [10] Nikolaos Ioannis Bountos, Arthur Ouaknine, and David Rolnick. Fomo-bench: a multi-modal, multi-scale and multi-task forest monitoring benchmark for remote sensing foundation models, 2024. 8
- [11] 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(7832):78–82, 2020. 2, 3
- [12] Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaeearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025. 5, 7
- [13] Boan Chen, Quanlong Feng, Bowen Niu, Fengqin Yan, Bingbo Gao, Jianyu Yang, Jianhua Gong, and Jiantao Liu. Multi-modal fusion of satellite and street-view images for urban village classification based on a dual-branch deep neural network. *International Journal of Applied Earth Observation and Geoinformation*, 109:102794, 2022. 8
- [14] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David B. Lobell, and Stefano Ermon. SatMAE: Pre-training transformers for temporal and multi-spectral satellite imagery. In *Advances in Neural Information Processing Systems*, 2022. 3, 8
- [15] Adrià Descals, Serge Wich, Erik Meijaard, David LA Gaveau, Stephen Peedell, and Zoltan Szantoi. High-resolution global map of smallholder and industrial closed-canopy oil palm plantations. *Earth System Science Data Discussions*, 2020:1–22, 2020. 2, 3, 4, 5, 7
- [16] A. Descals, S. Wich, Z. Szantoi, M. J. Struebig, R. Dennis, Z. Hatton, T. Ariffin, N. Unus, D. L. A. Gaveau, and E. Meijaard. High-resolution global map of closed-canopy coconut palm. *Earth System Science Data*, 15(9):3991–4010, 2023. 3, 7
- [17] European Commission. Regulation on deforestation-free products, 2023. Accessed: 2024-11-14. 1
- [18] Matthew E Fagan, Do-Hyung Kim, Wesley Settle, Lexie Ferry, Justin Drew, Haven Carlson, Joshua Slaughter, Joshua Schaferbien, Alexandra Tyukavina, Nancy L Harris, et al. The expansion of tree plantations across tropical biomes. *Nature Sustainability*, 5(8):681–688, 2022. 3, 4, 5, 7
- [19] Google Earth Engine. Satellite Embedding V1 — Earth Engine Data Catalog. https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL, 2025. 5
- [20] Jiaqi Gu, Hyoukjun Kwon, Dilin Wang, Wei Ye, Meng Li, Yu-Hsin Chen, Liangzhen Lai, Vikas Chandra, and David Z Pan. Multi-scale high-resolution vision transformer for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12094–12103, 2022. 2, 3, 8
- [21] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, et al. Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27672–27683, 2024. 8
- [22] Boran Han, Shuai Zhang, Xingjian Shi, and Markus Reichstein. Bridging remote sensors with multisensor geospatial foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27852–27862, 2024. 3
- [23] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J.

- Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853, 2013. 2, 4
- [24] Chia-Yu Hsu, Wenwen Li, and Sizhe Wang. Geospatial foundation models for image analysis: Evaluating and enhancing nasa-ibm prithvi’s domain adaptability. *International Journal of Geographical Information Science*, pages 1–30, 2024. 8
- [25] Amy Ickowitz, Stepha McMullin, Todd Rosenstock, Ian Dawson, Dominic Rowland, Bronwen Powell, Kai Mausch, Houria Djoudi, Terry Sunderland, Mulia Nurhasan, et al. Transforming food systems with trees and forests. *The Lancet Planetary Health*, 6(7):e632–e639, 2022. 1
- [26] Nikolai Kalischek, Nico Lang, Cécile Renier, Rodrigo Caye Daudt, Thomas Addoah, William Thompson, Wilma J Blaser-Hart, Rachael Garrett, Konrad Schindler, and Jan D Wegner. Cocoa plantations are associated with deforestation in côte d’ivoire and ghana. *Nature Food*, 4(5):384–393, 2023. 1, 3
- [27] Rudraksh Kapil, Seyed Mojtaba Marvasti-Zadeh, Nadir Erbilgin, and Nilanjan Ray. Shadowsense: Unsupervised domain adaptation and feature fusion for shadow-agnostic tree crown detection from rgb-thermal drone imagery. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 8266–8276, 2024. 3
- [28] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanus Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, pages 5637–5664. PMLR, 2021. 2
- [29] Alexandre Lacoste, Nils Lehmann, Pau Rodriguez, Evan Sherwin, Hannah Kerner, Björn Lütjens, Jeremy Irvin, David Dao, Hamed Alemohammad, Alexandre Drouin, et al. Geobench: Toward foundation models for earth monitoring. *Advances in Neural Information Processing Systems*, 36: 51080–51093, 2023. 3, 8
- [30] Yuting Liu, Liu Yang, and Yu Wang. Hierarchical fine-grained visual classification leveraging consistent hierarchical knowledge. In *Machine Learning and Knowledge Discovery in Databases. Research Track*, pages 279–295, Cham, 2024. Springer Nature Switzerland. 2
- [31] Robert N Masolele, Diego Marcos, Veronique De Sy, Itohan-Osa Abu, Jan Verbesselt, Johannes Reiche, and Martin Herold. Mapping the diversity of land uses following deforestation across africa. *Scientific Reports*, 14(1):1681, 2024. 1, 5
- [32] Matías Mendieta, Boran Han, Xingjian Shi, Yi Zhu, and Chen Chen. Towards geospatial foundation models via continual pretraining. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16806–16816, 2023. 8
- [33] DT Milodowski, ETA Mitchard, and Mathew Williams. Forest loss maps from regional satellite monitoring systematically underestimate deforestation in two rapidly changing parts of the amazon. *Environmental Research Letters*, 12(9): 094003, 2017. 2
- [34] Florencia Montagnini and Ruth Metzler. The contribution of agroforestry to sustainable development goal 2: end hunger, achieve food security and improved nutrition, and promote sustainable agriculture. In *Integrating landscapes: Agroforestry for biodiversity conservation and food sovereignty*, pages 21–67. Springer, 2024. 1
- [35] Kodoth Prabhakaran Nair. Tree crops. *Harvesting Cash from the World’s Important Cash Crops, 1st ed.*; Springer Nature: Cham, Switzerland, pages 249–285, 2021. 1
- [36] Vishal Nedungadi, Ankit Kariryaa, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. Mmearth: Exploring multi-modal pretext tasks for geospatial representation learning. In *European Conference on Computer Vision*, pages 164–182. Springer, 2024. 3, 8
- [37] Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, and Fahad Shahbaz Khan. Rethinking transformers pre-training for multi-spectral satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 27811–27819, 2024. 8
- [38] Paul Palmer, Caroline Wainwright, and Bo et al Dong. Drivers and impacts of eastern african rainfall variability. *Nature Reviews Earth Environment*, 4:254–270, 2023. 5
- [39] Luis Miguel Pazos-Outón, Cristina Nader Vasconcelos, Anton Raichuk, Anurag Arnab, Dan Morris, and Maxim Neumann. Planted: A dataset for planted forest identification from multi-satellite time series. In *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 7066–7070, 2024. 2, 3
- [40] Peter Potapov, Matthew C Hansen, Amy Pickens, Andres Hernandez-Serna, Alexandra Tyukavina, Svetlana Turubanova, Viviana Zalles, Xinyuan Li, Ahmad Khan, Fred Stolle, et al. The global 2000–2020 land cover and land use change dataset derived from the landsat archive: first results. *Frontiers in Remote Sensing*, 3:856903, 2022. 2, 4
- [41] Ravi Prabhu, Edmundo Barrios, Jules Bayala, Lucien Diby, Jason Donovan, Amos Gyau, Lars Graudal, R Jamnadass, J Kahia, K Kehlenbeck, et al. Agroforestry: realizing the promise of an agroecological approach. In *FAO. Agroecology for Food Security and Nutrition: Proceedings of the FAO International Symposium*, pages 201–224, 2015. 1
- [42] Colorado J Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Kurt Keutzer, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-mae: A scale-aware masked autoencoder for multiscale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4088–4099, 2023. 2, 3, 8
- [43] Zhiyuan Ren, Yiyang Su, and Xiaoming Liu. Chatgpt-powered hierarchical comparisons for image classification. In *Advances in Neural Information Processing Systems*, pages 69706–69718. Curran Associates, Inc., 2023. 2
- [44] Jessica Richter, Elizabeth Goldman, Nancy Harris, David Gibbs, Melissa Rose, Suzanne Peyer, Sarah Richardson, and Hemalatha Velappan. Spatial database of planted trees (sdpt version 2.0). 2024. 3, 4

- [45] John Wilson Rouse, Robert H. Haas, John A. Schell, and D. W. Deering. Monitoring vegetation systems in the great plains with erts. 1973. [5](#)
- [46] Masanobu Shimada, Takuya Itoh, Takeshi Motooka, Manabu Watanabe, Tomohiro Shiraishi, Rajesh Thapa, and Richard Lucas. New global forest/non-forest maps from alos palsar data (2007–2010). *Remote Sensing of Environment*, 155:13–31, 2014. [2](#), [4](#), [7](#), [8](#)
- [47] Maofeng Tang, Andrei Cozma, Konstantinos Georgiou, and Hairong Qi. Cross-scale mae: A tale of multiscale exploitation in remote sensing. *Advances in Neural Information Processing Systems*, 36:20054–20066, 2023. [3](#)
- [48] Jamie Tolan, Hung-I Yang, Benjamin Nosarzewski, Guillaume Couairon, Huy V Vo, John Brandt, Justine Spore, Sayantan Majumdar, Daniel Haziza, Janaki Vamaraju, et al. Very high resolution canopy height maps from rgb imagery using self-supervised vision transformer and convolutional decoder trained on aerial lidar. *Remote Sensing of Environment*, 300:113888, 2024. [5](#)
- [49] Robert Tropek, Ondřej Sedláček, Jan Beck, Petr Keil, Zuzana Musilová, Irena Šimová, and David Storch. Comment on “high-resolution global maps of 21st-century forest cover change”. *Science*, 344(6187):981–981, 2014. [2](#)
- [50] Angela Tsao, Ikenna Nzewi, Ayodeji Jayeoba, Uzoma Ayogu, and David B Lobell. Canopy height mapping for plantations in nigeria using gedi, landsat, and sentinel-2. *Remote Sensing*, 15(21):5162, 2023. [7](#), [8](#)
- [51] Héloïse Tschora and Francesco Cherubini. Co-benefits and trade-offs of agroforestry for climate change mitigation and other sustainability goals in west africa. *Global Ecology and Conservation*, 22:e00919, 2020. [1](#)
- [52] Jörg Wagner, Volker Fischer, Michael Herman, Sven Behnke, et al. Multispectral pedestrian detection using deep fusion convolutional neural networks. In *ESANN*, pages 509–514, 2016. [6](#)
- [53] Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M Albrecht, and Xiao Xiang Zhu. Ssl4eos12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, 2023. [3](#)
- [54] Yi Wang, Conrad M. Albrecht, and Xiao Xiang Zhu. Multi-label guided supervised contrastive learning for earth observation pretraining. In *IGARSS*, pages 7568–7571, 2024. [3](#), [6](#), [7](#)
- [55] Jonatas Wehrmann, Ricardo Cerri, and Rodrigo Barros. Hierarchical multi-label classification networks. In *International conference on machine learning*, pages 5075–5084. PMLR, 2018. [2](#)
- [56] Ben G Weinstein, Sergio Marconi, Stephanie A Bohlman, Alina Zare, Aditya Singh, Sarah J Graves, and Ethan P White. A remote sensing derived data set of 100 million individual tree crowns for the national ecological observatory network. *eLife*, 10:e62922, 2021. [3](#)
- [57] Global Forest Watch World Resources Institute. Spatial database of planted trees: Mapping planted forests on a global scale. [3](#)
- [58] Leikun Yin, Rahul Ghosh, Chenxi Lin, David Hale, Christoph Weigl, James Obarowski, Junxiong Zhou, Jessica Till, Xiaowei Jia, Nanshan You, Troy Mao, Vipin Kumar, and Zhenong Jin. Mapping smallholder cashew plantations to inform sustainable tree crop expansion in benin. *Remote Sensing of Environment*, 295:113695, 2023. [3](#)
- [59] Miao Zhang, Harvineet Singh, Lazarus Chok, and Rumi Chunara. Segmenting across places: The need for fair transfer learning with satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2916–2925, 2022. [2](#)
- [60] Shu Zhang, Ran Xu, Caiming Xiong, and Chetan Ramiah. Use all the labels: A hierarchical multi-label contrastive learning framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16660–16669, 2022. [2](#), [8](#)
- [61] Haotian Zhao, Justin Morgenroth, Grant Pearse, and Jan Schindler. A systematic review of individual tree crown detection and delineation with convolutional neural networks (cnn). *Current Forestry Reports*, 9(3):149–170, 2023. [3](#)