HarvestNet: A Dataset for Detecting Smallholder Farming Activity Using Harvest Piles and Remote Sensing

Jonathan Xu [*2], Amna Elmustafa [*1], Liya Weldegebriel [1], Emnet Negash [3,4], Richard Lee [1], Chenlin Meng [1], Stefano Ermon [1], David Lobell [1]

[1] Stanford University, [2] University of Waterloo, [3] Ghent University, [4] Mekelle University
contact@jonathanxu.com, {amna97, liyanet}@stanford.edu, emnet.negash@ugent.be,
rjlee6@stanford.edu, {chenlin, ermon}@cs.stanford.edu, dlobell@stanford.edu

Abstract

Small farms contribute to a large share of the productive land in developing countries. In regions such as sub-Saharan Africa, where 80% of farms are small (under 2 ha in size), the task of mapping smallholder cropland is an important part of tracking sustainability measures such as crop productivity. However, the visually diverse and nuanced appearance of small farms has limited the effectiveness of traditional approaches to cropland mapping. Here we introduce a new approach based on the detection of harvest piles characteristic of many smallholder systems throughout the world. We present HarvestNet, a dataset for mapping the presence of farms in the Ethiopian regions of Tigray and Amhara during 2020-2023, collected using expert knowledge and satellite images, totaling 7k hand-labeled images and 2k ground-collected labels. We also benchmark a set of baselines including SOTA models in remote sensing with our best models having around 80% classification performance on hand labelled data and 90%, 98% accuracy on ground truth data for Tigray, Amhara respectively. We also perform a visual comparison with a widely used pre-existing coverage map and show that our model detects an extra 56,621 hectares of cropland in Tigray. We conclude that remote sensing of harvest piles can contribute to more timely and accurate cropland assessments in food insecure regions.

Introduction

Smallholder farming is the most common form of agriculture worldwide, supporting the livelihoods of billions of people and producing more than half of food calories (Samberg et al. 2016; Lowder, Skoet, and Raney 2016). Cost effective and accurate mapping of farming activity can thus aid in monitoring food security, assessing impacts of natural and human-induced hazards, and informing agriculture extension and development policies. Yet smallholder farms are often sparse and fragmented which makes producing adequate and timely land use maps challenging, especially in resource constrained regions. Consequently, many land use datasets (Zanaga et al. 2022; Brown et al. 2022; Buchhorn et al. 2020) are inaccurate and updated infrequently in such regions, if at all.

Machine learning algorithms for remote sensing have proved to be successful in many sustainability-related measures such as poverty mapping, vegetation and crop mapping as well as health and education measures (Yeh et al. 2021). Moreover, satellite images are now widely available at different resolutions with global coverage at low to no cost (Planet Labs 2023). The performance of methods for mapping croplands in smallholder systems, however, remains limited in many cases (Zanaga et al. 2022; Brown et al. 2022).

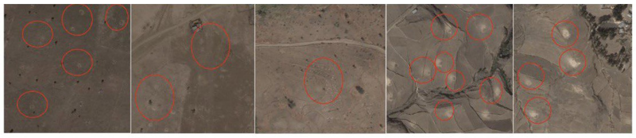


Figure 1: Various examples of harvest piles

Existing approaches to mapping croplands typically rely on either the unique temporal pattern of vegetation growth and senescence in crop fields compared to surrounding vegetation, the identification of field boundaries in high-resolution imagery, or some combination of both (Estes et al. 2022; Rufin et al. 2022). In non-mechanized smallholder systems like Ethiopia, where subsistence rain-fed agriculture predominates (Asmamaw 2017), these techniques face limitations. Weeds and wild vegetation often exhibit growth and spectral reflectance patterns resembling cultivated crops, causing confusion in spectral-based classification. The landscape's heterogeneity in smallholder systems, encompassing various land uses such as crops, fallow land, and natural vegetation, also poses challenges in accurately demarcating field boundaries and distinguishing different land cover types.

We highlight another feature that is common in smallholder systems throughout the world — the presence of harvest piles on or near fields that cultivate grains at the end of a harvest season. Crops, particularly grains, are manually cut and gathered into piles of 3-10m before threshing, a process of separating the grain from the straw. Figure 2 shows what a harvest pile can look like on a natural image scale. The harvest pile footprints are present until after threshing and finally disappear when the land is prepared for the upcoming season. Since the piles are valuable, they are not abandoned in fields. Unlike houses, roads, and field boundaries, harvest

[*] These authors contributed equally.

piles are a more dynamic indicator that signifies seasonal farming.

We focus our work on Ethiopia, which boasts the third largest agricultural sector in Africa based on its GDP (Statista 2021). Specifically, our attention is directed towards the lowlands in the Tigray and Amhara regions. This focus is driven by two main factors: Firstly, the area has historically been incorrectly mapped in previous works (Zanaga et al. 2022). Secondly, this area covers arid to subhumid tropical agroclimatic zones within Ethiopia, where we have available ground data. The major crops grown in these regions include teff, barley, wheat, maize, sorghum, finger millet, and sesame (ESS 2023; Sirany and Tadele 2022).

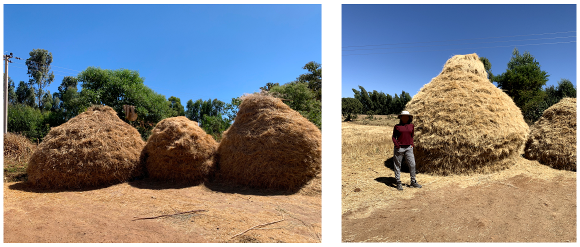


Figure 2: Photos of harvest piles. Left: person for scale.

Harvest pile detection is a novel task, thus we needed to hand label our dataset to train models. To gather labels for the presence of a pile in each image, we undertook a rigorous process of hand-labelling SkySat satellite images. In this process, experts - who are researchers originally from the region and have significant field and research experience in agricultural extension work in the region - guided the identification of key areas in Tigray and Amhara. Satellite images were obtained within these areas and then AWS Mturk identified the obvious negatives while experts labelled the positives. Figure 1 is a collection of various examples of piles in satellite images. In Figure 3, we show remote sensing examples of harvest piles at various stages of harvest. We then used this labelled data to train some SOTA models in remote sensing such as CNNs and transformers and achieved 80% accuracy on the best model. Moreover, we generated a map depicting projected farming activities in Tigray and Amhara regions, and compared it with the most current cover map.

Our contributions are as follows:

- We propose a framework to detect farming activity through the presence of harvest piles.
- We introduce HarvestNet, a dataset of around 7k satellite images labeled by a set of experts collected for Tigray and Amhara regions of Ethiopia around the harvest season of 2020-2023.
- We document a multi-tiered data labelling pipeline to achieve the optimal balance of scale, quality, and consistency.
- We benchmarked SOTA models on HarvestNet and tested them against ground truth data and hand-labeled data to show their efficacy for the task.
- We produced a map for the predicted farming activity by running inference on the unlabelled data, and compared it against ESA WorldCover (Zanaga et al. 2022), the most updated land usage cover map.

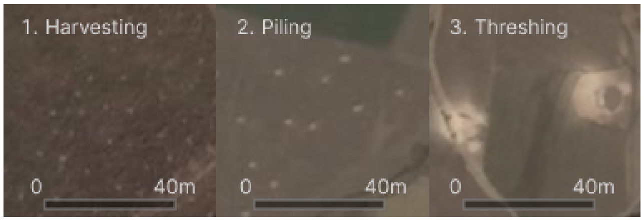


Figure 3: Various stages of harvest activity

Related Work

Mapping croplands using remote sensing has been well researched in the past (Kussul et al. 2017; Jiang et al. 2020; Friedl et al. 2002; Zanaga et al. 2022; Buchhorn et al. 2020; Brown et al. 2022; Kerner et al. 2020). Some methods use feature engineering with nonlinear classifiers (Zanaga et al. 2022; Jiang et al. 2020; Brown et al. 2022), others use deep learning methods (Kerner et al. 2020; Kussul et al. 2017). In all these works, the Normalized Difference Vegetation Index (NDVI) as well as multispectral satellite bands are used as an input, NDVI is a numerical indicator used to quantify the presence and vigor of live green vegetation by measuring the difference between the reflectance of near-infrared (NIR) and visible (red) light wavelengths in imagery. ESA (Zanaga et al. 2022) and Dynamic World (Brown et al. 2022) combine both NDVI and multispectral bands to provide global coverage of more than 10 classes of land use, which include crop coverage. These maps are the largest in scale and have a pixel resolution of 10m. Other methods (Mananze, Pôças, and Cunha 2020; Hackman, Gong, and Wang 2017; Kerner et al. 2020; Acharki 2022) introduced a higher resolution but on a smaller scale in countries such as Mozambique, Ghana, Togo and Morocco.

Active learning is a method of building efficient training sets by iteratively improving the model performance through sampling. Some studies (Estes et al. 2022; Rufin et al. 2022) have employed active learning to map smallholder farms. This approach helps mitigate bias in cropland mapping, as it can more accurately detect larger fields compared to other methods. However, none of these works have explored the concept of utilizing harvest piles as indicators when mapping smallholder farms.

Method

In many smallholder farms for crops such as grains, farmers collect the harvest into piles during the harvest season, in preparation for threshing. These piles can be heaps of various crop types gathered around the nearest threshing ground. Therefore, the detection of piles during the harvest season is a very compelling indicator of farming activity. We suggest employing RGB satellite imagery for pile detection, as these color bands are widely accessible and can be easily and economically adjusted for other purposes beyond pile detection.

Task formulation

To provide a proof of concept of this novel method, we defined the task of farmland detection as a binary classification task where the input is a square RGB satellite image at a predefined scale. If l is a location represented by latitude and longitude, the task is to build a machine learning model that takes a satellite image x_l and predicts y_l where y_l is a binary output indicating the presence of farming activity at location l. The output should be positive if the image contains at least one indication of harvest activity. In our area of interest, which covers Tigray and Amhara regions in Ethiopia, the harvest process constists of three stages: cutting down and grouping crops to be collected (harvesting; Figure 3 left), piling the crops to be processed (piling; Figure 3 middle), and processing the piles to separate grains from the straw (threshing; Figure 3 right). Each stage results in different footprints of harvest patches. We classify the presence of any of these stages as a positive example of harvest activity and we use binary cross entropy loss defined by

$$L_{CE} = \frac{1}{N} \sum_{l} -y_l \cdot \log(\hat{y}_l) - (1 - y_l) \log(1 - \hat{y}_l) \quad (1)$$

where N is the number of locations l, y_l the predictions and $\hat{y}_l$ the ground truth presence of harvest piles. More examples of harvest piles are displayed in Appendix Figure A3 and A2.

HarvestNet Dataset

Here we introduce HarvestNet, the first dataset to our knowledge created for the task of detecting harvest activity from pile detection. Ethiopia is the second most populated country in the continent, with a majority of its people primarily dependent on smallholder rain-fed agriculture. In our regions of interest, the piling of harvests occurs during Meher, the main harvest season between September and February. These piles can be observed as early as October and stay on the land as late as May of the next year. We therefore restrict the time samples of our dataset to Oct-May months. A geographical scale of around 250 m was found to be a good fit for our purposes since piles are typically located within 1km from the field plot. Our images thus cover square land areas of dimensions 256x256 m.

Satellite images We use images from 2 different resolutions (0.5m per pixel and 4.77m per pixel). This is due to the small size of the harvest piles in an image, which makes hand labeling, as well as more accurate mapping only possible on a higher-resolution image, as shown in Figure 4.

On the other hand, higher-res images are limited in coverage and availability. Thus, we also include around 9k (7k labeled images + 2k ground truth images) lower-res images as part of our dataset. We use the high-res images (150k unlabelled, 7k labeled images) for training and testing on the hand-labeled test set as well as for creating the crop map, while we use the low-res images for the ground truth testing since the higher res is not available in the ground truth locations. In our dataset, each example consists of a unique latitude, longitude, altitude, and date. All examples correspond to a SkySat image, and all labelled examples correspond to both a SkySat and a PlanetScope image.

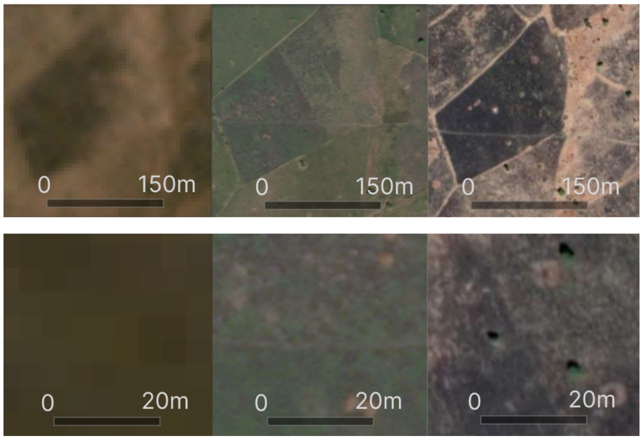


Figure 4: Side by side comparison of two areas, captured in 4.77m (left), 0.5m (center) and 0.3m (right) resolution. Note that piles become indistinguishable at 4.77m resolution.

SkySat images (planet labs 2023) are 512x512 pixel subsets of orthorectified composites of SkySat Collect captures at a 0.50 meter per pixel resolution. SkySat images are normalized to account for different latitudes and times of acquisition, and then sharpened and color corrected for the best visual performance. For our analysis, we downloaded every SkySat Collect with less than 10 percent cloud cover between October 2022 and January 2023. In total, we have 157k SkySat images, of which 7k are labeled.

PlanetScope images (Planet Labs 2023) are subsets of monthly PlanetScope Visual Basemaps with a resolution of 4.77 meters per pixel. These base maps are created using Planet Lab's proprietary ”best scene on top” algorithm to select the highest quality imagery from Planet's catalog over specified time intervals, based on cloud cover and image sharpness. The images include red, green, blue, and alpha bands. The alpha mask indicates pixels where there is no data available. We used subsets that correspond to the exact location and month of each of the 7k hand-labelled SkySat images. To maintain the same coverage of 256x256 m at the lower resolution, we used the bounding box of each SkySat image to download PlanetScope images at a size of roughly 56x56 pixels. Since the PlanetScope images are readily available and have good coverage in geography and time series, we separately downloaded 4 PlanetScope images for each area of interest corresponding to the 2k ground truth images collected by the survey team. They include a capture for each month in the Oct-Jan harvest season. This window guarantees that farming activity will be captured in at least one of the 4 images.

Labelling Since this is a novel task, we hand-labeled our entire training and test set. We wanted to create a high-quality, high-coverage dataset despite having limited resources and sparse access to field data and subject experts familiar with remote sensing on harvest piles. Thus, we developed a multi-staged committee approach to label successively more focused data sets. With the guidance of experts in the agricultural zones of our region of interest, we drew polygons around different areas distributed around Tigray and Amhara, making sure that these areas had at least a 2:1 negative to positive class ratio. We then filtered out obvious negatives such as mountains, and shrubs using crowd-sourced labelling powered by Amazon Mechanical Turk. When there is a disagreement between the MTurk labellers, we made the decision of whether there is potential farming activity in the image. The potential positive examples were then given to experts, who hand-labeled whether the image contained actual harvest activity. In Appendix Figure A4 we outline our labelling process in greater detail. The labelling process was done through inspection on SkySat images exclusively, afterwards PlanetScope images were paired with the corresponding labelled SkySat images. By the end of this stage, we had roughly 7k labelled examples, which each consisted of a SkySat image of size 512x512 pixels and a PlanetScope image of size 56x56 pixels covering the same area at the same month.

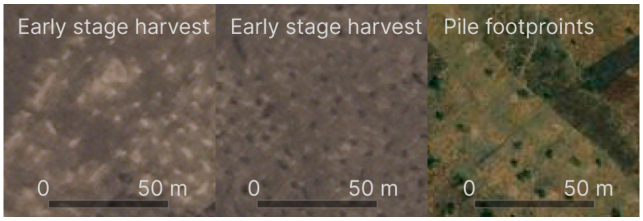


Figure 5: Examples of harvest pile activity that are not strictly piles.

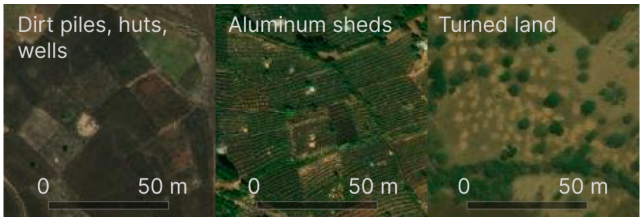


Figure 6: Examples of edge cases that are not harvest piles.

During the labeling process, we encountered diverse edge cases. Some image features resulted from the harvest piling process but did not match the conventional stage of harvest activity shown in Figure 1. Notable examples, depicted in Figure 5, include early-stage light and dark crop bunches and residual pile footprints. These were labeled as positive instances. Additionally, some images depicted small dots resembling harvest piles, which were later identified, through consultation with our experts, as various entities such as dirt piles, aluminum sheds, and altered land shown in Figure 6. These were deemed unrelated to harvest activity and marked as negative instances.

Ground Truth In March 2023, we sent a survey team to gather ground truth data in specified locations in Tigray and Amhara, in order to validate the prediction of our models for the 2022-2023 harvest season. 1,017 and 1,279 labels were gathered in Tigray and Amhara regions respectively. Ground truth data were gathered for all harvest crop types, including maize, teff, wheat, and finger millet. All the heaps belong to the pile point category and are situated within a maximum distance of 500 meters from the field plot. A map of ground truth collection zones is plotted in Appendix Figure A5. Due to the ongoing armed conflict, the team was unable to visit areas in Tigray that were covered by SkySat (higher-res imagery) in our image dataset. In response, we opted to combine the ground truth data with PlanetScope images, a more diverse collection that spans the geographic area with an extensive temporal range.

Dataset split As the goal is to build a dataset that is well-balanced, we aimed for a roughly equal split of positive and negative labels. We were able to collect SkySat images from various regions shown in Figure 7, that are representative of the diversity of the geography. The exact distribution of the dataset geography and labels is described in Appendix Figure A1.

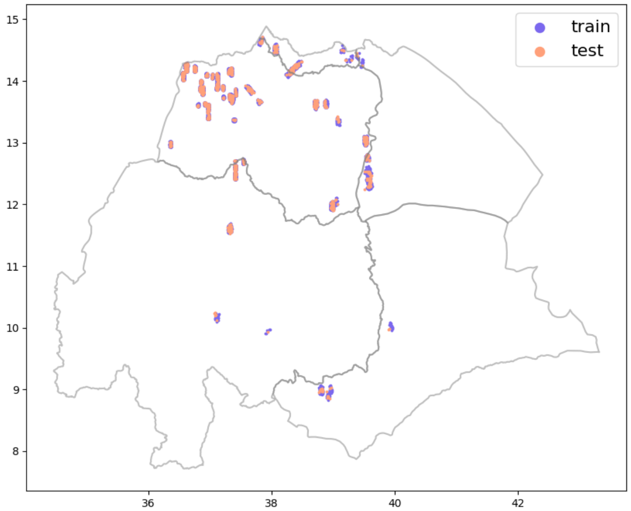


Figure 7: Train-Test split

Since the image captures are distributed spatially, it is desirable to avoid image overlap between the training and test sets while ensuring a roughly similar distribution between sets. Thus, we used a graph traversal approach by creating vertices for each image's coordinates, and edges connecting images that overlap. We first identified the connected components of the shape graphs as shown in Figure 8 A (the code is provided in Appendix Listing A1). Afterwards, the partitions were distributed evenly between the train and test set as shown in Figure 8 B. This was carried out by moving the two largest partition groups to the train set, and then iteratively moving the remaining groups to either the train or test set to maintain an 80:20 ratio between the train and test sets. This results in a train/test split that strictly do not overlap geographically, while still sharing a similar geographic distribution.

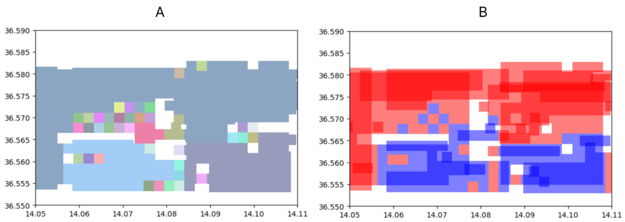


Figure 8: (A): A region of image captures, organized into non-overlapping partitions of overlapping shapes. All partitions are assigned a random color. (B): The partitions are then divided into train (red) and test (blue) that do not overlap with each other.

Benchmarking

We trained various machine learning models on our dataset to predict the presence of harvest activity in an image, as described below.

MOSAIKS (Rolf et al. 2021) This approach employs a non-deep learning technique to extract features from a satellite image. It achieves this by convolving a series of randomly chosen patches with the input image. Subsequently, these extracted features are used in various downstream tasks. The method offers good perform well in these tasks at a low cost. We featurized our dataset to contain 512 features per image, and then use an XGBoost classifier to predict our target from these features.

SATMAE (Cong et al. 2022) is a pretraining framework based on masked autoencoders (MAE). It is pretrained on the FMOW dataset (0.3 m resolution imagery) for single image and temporal tasks, and pretrained on images from Sentinel 2 (10 m) for multispectral tasks. The model demonstrated good performance on different downstream and transfer learning tasks. We employ transfer learning by training the model - pretrained on FMOW dataset - on our specific dataset to predict the presence of harvest piles.

Swin Autoencoder (Liu et al. 2022) is a type of vision transformer that builds hierarchical feature maps by merging image patches in deeper layers and has linear computation complexity to input image size by computing self-attention only within each local window. We pretrain a masked image autoencoder built on Swin Transformer V2, using our 150k Skysat images. The input image is scaled to 224x224 pixels, and divided into a grid of patches of size 28x28. We use a mask ratio of 40%. Then we attach a fully connected layer to the transformer's pooled output of dimension 1 x 768. The model is then fine tuned on our training set of labelled Skysat images.

Satlas (Bastani et al. 2022) is a pre-trained model based on the Swin transformer, and pretrained on 1.3 million remote sensing images collected from different sources. The model performs well for in-distribution and out of distribution tasks, suggesting the benefit of pretraining on a large dataset. We used the weights pretrained on higher res images, froze the model, and trained a fully connected layer on top of the pre-trained model.

ResNet-50 (He et al. 2016) Convolutional Neural Networks (CNNs) have proven to perform well in several remote sensing tasks. Here, we used ResNet-50, one of the most popular and efficient networks, to predict our target. Since our input satellite image is in RGB, we used the ImageNet initialization of the network and trained a supervised binary classification task using our labelled dataset.

Experiments

Experimental details

As our working dimensions are areas of size 256x256 m, we center cropped the SkySat images to 512x512 pixels and PlanetScope images to 56x56 pixels before normalizing to zero mean and unit standard deviation. These images were then scaled to fit the default input dimensions of the models.

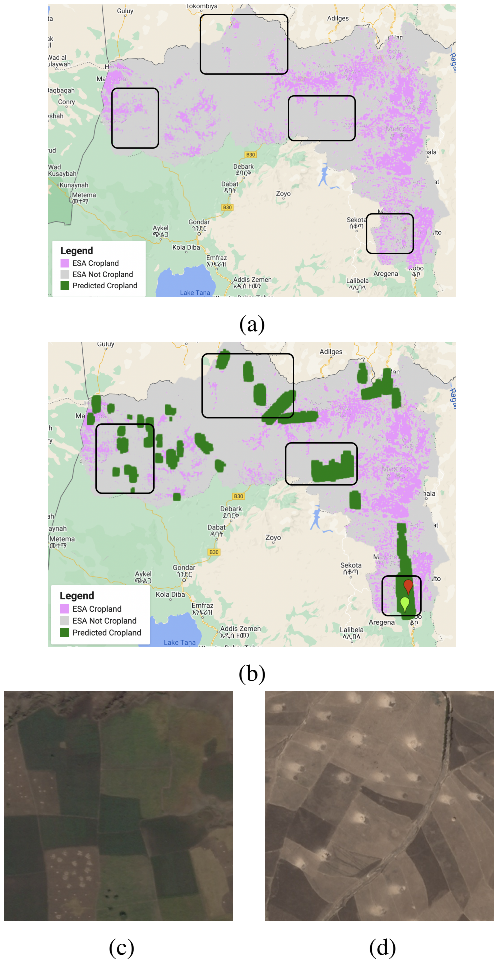
MOSAIKS was trained with 512 features. The deep models were trained using the Adam optimizer to minimize the binary cross-entropy loss criterion. The hyperparameters on batch size, learning rate, scheduler, and training step count are described in Appendix Table A1. We experimented with combinations of hyperparameters and settled on the best performing combinations. For transformers-based models we chose the batch size that would maximise use of the 24GB of VRAM in our graphics cards. The models were trained until they converged, and the step counts were recorded.

Evaluation

As the task of harvest pile detection depends on the nuances of real farm activity, it is always desirable to have both a qualitative test as well as a quantitative one. We describe both evaluations below.

Qualitative evaluation We compare the coverage of our predictions against ESA (Zanaga et al. 2022). ESA is a land use map, providing global coverage for 2020 and 2021 at 10 m resolution, developed and validated based on Sentinel-1 and Sentinel-2 data. It has been independently validated with a global overall accuracy of about 75%. Despite being SOTA in mapping land cover and land use, our experts identified many errors in smallholder systems as highlighted in Figure 9a. The pink overlay describe ESA's classification of land as cropland. In the squares outlined in the 9a, ESA fails to detect much of the farmland that is actually there. We map these locations using our best performing model and present a visual comparison of the two maps. We also provide visual confirmation of farm activity by visually referencing the input satellite images in two of these locations.

Quantitative evaluation In this evaluation, we calculate the classification performance of our trained models using accuracy, AUROC, precision and recall. We also use the same metrics to measure the performance of our models against ground truth data.



(a)

(b)

(c)

(d)

Figure 9: (a) The ESA map for our study region. (b) Positive predictions from our ResNet-50 model, overlaid the ESA map. Sampled locations are shown in squares, (c) and (d) show satellite images of the locations pinpointed in (b)

Results

Model	Accuracy	AUROC	Precision	Recall	F1-Score
Satlas	67.17	62.47	80.0	30.61	44.28
SatMAE	60.0	56.35	57.37	29.73	39.17
MOSAIKS	55.46	51.81	47.65	23.59	31.56
Swin Autoencoder	**80.87**	**80.15**	79.88	**74.79**	**77.23**
ResNet-50	79.18	77.85	**81.4**	67.61	73.87

Table 1: Results for the proposed models on the hand labelled test set

Model	Region	Accuracy	F1-Score	Precision	Recall
ResNet-50	Amhara	98.68	99.33	1	98.68
ResNet-50	Tigray	90.76	95.16	1	90.76

Table 2: Results for the ResNet model evaluated on the test ground truth data

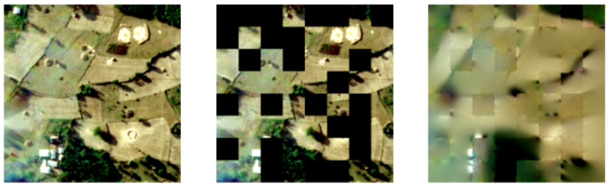


Figure 10: Reconstruction results from the Swin v2 masked autoencoder trained on 150k unlabelled Skysat imagery

	Total	- , +	+ , +	- , -
Number of samples	150577	11563	24076	38989
Area covered (ha)	986,821	56,621	62,082	137,059

Table 3: Results for the comparison between the positives and negatives predicted by the ResNet model and ESA. "-,+" are areas where ESA predicts negative while our model predicts positive (newly detected cropland), "+,+" and "-,-" are areas where our model and ESA agree

Table 1 displays our benchmark outcomes obtained from the HarvestNet dataset, utilizing the hand-labeled data as the test set. In Table 2, we show the results of the ResNet model on the ground truth data. We use the ResNet model here because it is one of the best models we have and also the best performing model in terms of precision.

Figure 10 demonstrates the reconstruction results of a self-supervised Swin masked autoencoder pretrained on the 150k SkySat dataset. We can see that although the model was not trained on the input image, it generalizes well on filling in the masked area for Ethiopian landscapes. The model was trained on an 80% split of the images, and evaluated on the remaining 20% split.

In Figure 9, a comparison is presented between the ESA map (Figure 9a) and our predicted map (Figure 9b). We emphasize particular regions outlined in black rectangles, wherein our experts have pinpointed inaccurately classified regions by ESA. A closer comparison of these regions of interest can be viewed in Appendix Figure A6. Satellite images for two sampled locations (Figure 9c and 9d) reveal specific examples where ESA fails to detect cropland accurately. Table 3 tabulates the variations in positive and negative predictions between our map coverage and the ESA map. The objective is to illustrate the number of samples we predict as positive or negative compared to ESA, along with instances where we predict positively while ESA predicts negatively. Additionally, we present the additional cropland area detected by our model (in hectares) and the overlapping cropland region shared between the two models.

Discussion

Model performance The outcomes presented in Table 1 highlight a notable trend: deep models consistently outperform non-deep models that rely on feature generators such as MOSAIKS. This disparity in performance can be attributed to the nature of our task, which involves identifying piles – intricate and compact elements within an image. The intricate nature of piles demands the capabilities of a deep network to adequately capture and detect these distinctive features.

The second notable finding is that ResNet-50 performs quite well, even outperforming SatMAE. We believe that this is mainly due to the fact that CNNs better maintain pixel structure and generate feature maps that retain spatial information, which is a critical aspect for accurately detecting piles. Following this trend, the Swin autoencoder slightly outperforms ResNet, thanks to its incorporation of low-level details in its hierarchical feature maps. Even though Satlas is based on the Swin architecture, it performs worse than the Swin autoencoder pretrained on our unlabelled images. This suggests that although it was pretrained on a very comprehensive dataset, it is not uniquely positioned to perform in specific areas of interest.

Lastly, it is worth noting that our models' precision are notably higher than their recall, indicating a relatively high dataset quality. However, to enhance performance, further inclusion of positive samples is required, presenting a potential avenue for future research.

Evaluation and coverage In Table 2, we observe notably promising outcomes for ResNet predictions on ground truth data. This relatively elevated performance, when compared to the results from hand-labeled test data, can be attributed to two main factors. Firstly, during ground truth testing, we employed four distinct images for each location, each representing a different month within the harvest season. By taking the union of the predictions from each image, we effectively enhance the prediction quality. This approach was feasible with the ground truth test set due to the utilization of PlanetScope imagery, which offers greater availability compared to SkySat but at a lower resolution. Secondly, the ground truth data exclusively comprises positive samples, meaning there are no inherent false positives by default. This inherent nature of the ground truth dataset significantly contributes to the elevated accuracy observed in our results.

In Table 3 and Figure 9, we aim to show that our method can improve upon ESA predictions. We show in Figure 9 that there are sampled locations (in rectangles) that ESA predicts as non-crop lands while our model predicts as cropland shown in figure 9b. Moreover, Figure 9c and 9d show 2 samples of satellite images corresponding to 2 of these locations. In table 3, we also show that there are around 11k examples (corresponding to 57k ha) where our model predicts farming activity and ESA does not, and around 62k samples (199k ha) where our model predicts the same as ESA. This added cropland is roughly estimated by experts to be 90% true cropland. This demonstrates the potential for using harvest pile features to improve existing maps in smallholder regions. In Appendix Figure A6 we show a higher zoom map for these missed locations.

Limitations and future work Due to resource limitations and the nature of small feature classification, our dataset has some limitations. We created binary labels on fixed 256x256 m areas, which resulted in a relatively low spatial resolution of the dataset. While this resolution was chosen to strike a practical balance in terms of land coverage, it is worth noting that one could break down the existing images into smaller subdivisions and conduct binary classification on subsections of positively labelled images. An important advantage of this approach is that all subdivisions derived from negatively labeled images in the HarvestNet dataset can serve as negative labels for training.

This dataset is also made for a binary classification task rather than object detection or semantic segmentation of harvest piles. The object detection approach could be helpful by giving information about the location, size, and density of piles. As a large part of the initial challenge was to identify areas that contain any piles at all, this next step can be applied to our existing positively labelled images. It may be a fruitful endeavor to explore using zero-shot image segmentation models such as Segment Anything (Kirillov et al. 2023) to automate the process.

As in Figure 3, Figure 5 and Appendix Figure A3, we see that there are various image features that correspond to harvest pile activity. Whether they are crops gathered for harvest, piles of harvest, threshing products, or footprints left by piles, they have all been classified as positive examples of harvest activity. When access to expert feedback becomes more readily available, it would be better to also classify each feature as a specific type of harvest activity. This improvement would complement the object detection direction, especially with images that have multiple types of harvest activity.

Another possible next step is to incorporate time series data to the detection of harvest piles. With this project, we faced manpower limitations to effectively hand-label a very large dataset, so we focused more on geographical diversity with our images. The performance of PlanetScope imagery to detect true positives suggests there are greater improvements to be made by training and inferencing on multiple time-series captures of the same area.

The performance of models trained on HarvestNet also suggests that similar approaches to targeting small but important image features can yield benefits in other agricultural settings, such as hay bale detection in North America. A dataset of harvest piles may offer opportunities for transfer learning to these other domains.

Conclusion

In this work, we present HarvestNet, the first dataset for detecting farming activity using remote sensing and harvest piles. HarvestNet includes a dataset for both Tigray and Amhara regions in Ethiopia, totaling 7k labelled SkySat images, and 9k labelled PlanetScope images corresponding to 2k ground truth points and the 7k labelled Skysat images. We document the process of building the dataset, present different benchmarks results on some of the SOTA remote sensing models, and conduct land coverage analysis by comparing our predictions to ESA, a SOTA land use map. We show in our comparison that we greatly improve the current ESA map by incorporating our method of pile detection. Thus, by combining our approach with existing coverage maps like ESA, we can have a direct impact on efforts to map active smallholder farming, consequently helping to better monitor food security, assess the impacts of natural and human-induced disasters, and inform agricultural extension and development policies. Link to the labels and benchmark code will be available in the supplementary material.

References

Acharki, S. 2022. PlanetScope contributions compared to Sentinel-2, and Landsat-8 for LULC mapping. *Remote Sensing Applications: Society and Environment*, 27: 100774.

Asmamaw, D. K. 2017. A critical review of the water balance and agronomic effects of conservation tillage under rain-fed agriculture in Ethiopia. *Land Degradation & Development*, 28(3): 843–855.

Bastani, F.; Wolters, P.; Gupta, R.; Ferdinando, J.; and Kembhavi, A. 2022. Satlas: A large-scale, multi-task dataset for remote sensing image understanding. *arXiv preprint arXiv:2211.15660*.

Brown, C. F.; Brumby, S. P.; Guzder-Williams, B.; Birch, T.; Hyde, S. B.; Mazzariello, J.; Czerwinski, W.; Pasquarella, V. J.; Haertel, R.; Ilyushchenko, S.; et al. 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1): 251.

Buchhorn, M.; Lesiv, M.; Tsendbazar, N.-E.; Herold, M.; Bertels, L.; and Smets, B. 2020. Copernicus global land cover layers—collection 2. *Remote Sensing*, 12(6): 1044.

Cong, Y.; Khanna, S.; Meng, C.; Liu, P.; Rozi, E.; He, Y.; Burke, M.; Lobell, D.; and Ermon, S. 2022. Satmae: Pretraining transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35: 197–211.

ESS. 2023. Agricultural Sample Survey, Report on area and production of major crops (2014 to 2015). Technical report, The Federal Democratic Republic of Ethiopia, Central Statistical Agency Ethiopian Statistics Service.

Estes, L. D.; Ye, S.; Song, L.; Luo, B.; Eastman, J. R.; Meng, Z.; Zhang, Q.; McRitchie, D.; Debats, S. R.; Muhando, J.; et al. 2022. High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales. *Frontiers in artificial intelligence*, 4: 744863.

Friedl, M. A.; McIver, D. K.; Hodges, J. C.; Zhang, X. Y.; Muchoney, D.; Strahler, A. H.; Woodcock, C. E.; Gopal, S.; Schneider, A.; Cooper, A.; et al. 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote sensing of Environment*, 83(1-2): 287–302.

Hackman, K. O.; Gong, P.; and Wang, J. 2017. New landcover maps of Ghana for 2015 using Landsat 8 and three popular classifiers for biodiversity assessment. *International Journal of Remote Sensing*, 38(14): 4008–4021.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Jiang, Y.; Lu, Z.; Li, S.; Lei, Y.; Chu, Q.; Yin, X.; and Chen, F. 2020. Large-scale and high-resolution crop mapping in china using sentinel-2 satellite imagery. *Agriculture*, 10(10): 433.

Kerner, H.; Tseng, G.; Becker-Reshef, I.; Nakalembe, C.; Barker, B.; Munshell, B.; Paliyam, M.; and Hosseini, M. 2020. Rapid response crop maps in data sparse regions. *arXiv preprint arXiv:2006.16866*.

Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A. C.; Lo, W.-Y.; et al. 2023. Segment anything. *arXiv preprint arXiv:2304.02643*.

Kussul, N.; Lavreniuk, M.; Skakun, S.; and Shelestov, A. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5): 778–782.

Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. 2022. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12009–12019.

Lowder, S. K.; Skoet, J.; and Raney, T. 2016. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World development*, 87: 16–29.

Mananze, S.; Pôças, I.; and Cunha, M. 2020. Mapping and assessing the dynamics of shifting agricultural landscapes using google earth engine cloud computing, a case study in Mozambique. *Remote Sensing*, 12(8): 1279.

Planet Labs. 2023. planet visual maps. https://developers.planet.com/docs/data/visual-basemaps/. Accessed: 2023-08-09.

planet labs. 2023. skysat docs. https://developers.planet.com/docs/data/skysat/. Accessed: 2023-08-09.

Rolf, E.; Proctor, J.; Carleton, T.; Bolliger, I.; Shankar, V.; Ishihara, M.; Recht, B.; and Hsiang, S. 2021. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1): 4392.

Rufin, P.; Bey, A.; Picoli, M.; and Meyfroidt, P. 2022. Large-area mapping of active cropland and short-term fallows in smallholder landscapes using PlanetScope data. *International Journal of Applied Earth Observation and Geoinformation*, 112: 102937.

Samberg, L. H.; Gerber, J. S.; Ramankutty, N.; Herrero, M.; and West, P. C. 2016. Subnational distribution of average farm size and smallholder contributions to global food production. *Environmental Research Letters*, 11(12): 124010.

Sirany, T.; and Tadele, E. 2022. Economics of Sesame and Its Use Dynamics in Ethiopia. *The Scientific World Journal*, 2022.

Statista. 2021. Contribution of agriculture, forestry, and fishing sector to the GDP in Africa as of 2021. https://www.statista.com/statistics/1265139/agriculture-as-a-share-of-gdp-in-africa-by-country/. Accessed: 2023-08-09.

Yeh, C.; Meng, C.; Wang, S.; Driscoll, A.; Rozi, E.; Liu, P.; Lee, J.; Burke, M.; Lobell, D. B.; and Ermon, S. 2021. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*.

Zanaga, D.; Van De Kerchove, R.; Daems, D.; De Keersmaecker, W.; Brockmann, C.; Kirches, G.; Wevers, J.; Cartus, O.; Santoro, M.; Fritz, S.; et al. 2022. ESA WorldCover 10 m 2021 v200.

Appendix

Image collection

All SkySat images were downloaded using the Planet Python SDK. This process includes account authentication, creating a session to call Planet servers, creating an order request, and downloading the order when it's ready. For our analysis we downloaded SkySat Collects, which are approximately 50-70 SkySat Scenes and 20 x 5.9 square kilometers in size. Collects were subsetted differently based on their use case. Images used for inference were produced by subsetting entire Collects into 512 x 512 pixel sized areas. Images that were partially empty were thrown away. Unlike PlanetScope, SkySat has very limited spatial and temporal availability, limiting our choices to specific regions of Tigray and Amhara. We addressed this issue while maintaining our quota by diversely sampling areas in Tigray and Amhara. All of our images were originally stored in different folders in Google Drive based on region and time, but were later merged into one folder while still maintaining temporal and spatial information.

Accessing the dataset

The dataset is made partially accessible through this link https://figshare.com/s/45a7b45556b90a9a11d2. The labels and PlanetScope images will be shared, but unfortunately we cannot release the SkySat images due to Planet Labs' licensing requirements which would render the labels useless. Additionally, the benchmark code can be found on GitHub: https://anonymous.4open.science/r/harvest-piles-9D64.

We provide the dataset in a .zip folder structured as follows:

```
Dataset
    |- planetscope_images/
    |- lables_all.csv
    |- train.csv
    |- test.csv
```

Computational resources

We trained our models on a single NVIDIA GeForce RTX 2080 Ti GPU with a fixed seed. MOSAIKS was trained with 3 different seeds and the average of these seeds was reported. The Swin masked autoencoder was pretrained on the task of reconstructing masked patches, and the model converged in 23 hours. The pretrained models were fine tuned for at most 5 hours.

Training parameters

In Appendix Table A1, we outline the different hyperparameters of the deep models we used. Our models were all trained for 200 epochs, and the epoch count where they converged is recorded in the table. All other unlisted parameters were set to their defaults.

Table A1: Hyperparameters of models trained on HarvestNet

Model	Batch size	Scheduler	Learning rate	Training steps	Convergence epochs
Satlas	50	Warmup cosine	3e-4	6000	55
SatMAE	64	Warmup cosine	3e-4	2500	29
Swin Autoencoder	50	Linear	1e-3	4500	40
ResNet-50	32	One cycle	1e-3	2600	15

Split counts

In Appendix Table A2, we provide counts for each train test split in both Tigray and Amhara, we also show counts of positives and negative examples in each split.

Table A2: Split counts for the train and test set, based on region and label

	Tigray	Amhara	Positives	Negatives	Total
Train	4737	795	2547	2985	5532
Test	1171	212	608	781	1383

Ablation studies

In this section we explore the impact of various hyperparameters on the performance of models trained on HarvestNet.

Table A3: ResNet-50 Ablations

Pretrain	Optimizer	Accuracy	AUROC	Precision	Recall	F1-Score
None	Adam	65.05	69.31	60.00	60.41	59.34
IMAGENET1K_V2	Adam	79.18	87.23	79.04	71.75	74.19
IMAGENET1K_V2	MADGRAD	**79.85**	**88.45**	**80.34**	**72.65**	**75.65**

ResNet-50 was trained using fp16 mixed precision, using the one_cycle_lr learning rate scheduler with a learning rate of 0.001.

Table A4: Satlas ablations

Variation	Accuracy	AUROC	Precision	Recall	F1-Score
Modify pretrained output layer	64.78	60.08	**82.76**	24.00	37.21
Append new output layer	**67.17**	**62.47**	80.0	**30.61**	**44.28**

We first modified the default Satlas model by modifying its final projection layer output dimension from 1000 to 1, and appending a sigmoid layer on top.

We then modified the default Satlas model by appending an FC layer with input dimension 1000 and output dimension 1 to the model, and appending a sigmoid layer on top. This performed better, which we believe is due to the fact that appending a layer maintains of the the latents learned in the pretrained weights.

Table A5: Swin ablations

Freeze pretrained	Accuracy	AUROC	Precision	Recall	F1-Score
Yes	70.10	68.57	68.28	57.43	62.37
No	**80.87**	**80.15**	**79.88**	**74.79**	**77.23**

Dataset distribution

In Appendix Figure A1, we show distributions of latitude, longitude and altitude on train, test sets as well as on the entire labelled set and unlabelled set. One notable feature of the dataset is that for each bucket in the histogram, there is a roughly equal number of positive and negative labels. Moreover, the ratio of train to test is also around 80:20 in all buckets. Most of our labelled altitude was between 500-1000m, this is because we were targeting lowlands, since previous work (Zanaga et al. 2022) had errors in lowlands in particular.

Examples of harvest piles

In Appendix Figure A2 and Appendix Figure A3 we provide more examples of harvest activity.

Ground truth collection

During February and March 2023, we sent teams of six individuals to Tigray and Amhara regions respectively to collect ground truth data. These teams had diverse backgrounds: Tigray's team included staff from Mekelle University's Department of Dryland Crop and Horticultural Sciences and Department of Land Resources Management and Environmental Protection, and staff from the College of Agriculture and Natural Resources in Mekelle, Tigray. The Amhara team was comprised of staff from the Irrigation and Lowland Area Development Bureau in Bahir Dar, Amhara.

To gather data, the teams used handheld GPS devices, rental cars, pens, notebooks, and laptops for encoding. Guided by a map featuring available SkySat images in the 2022 harvest season, the team selected sites near roads for accessibility. Local farmers played a vital role in locating harvest pile sites. Importantly, no gathered data was discarded throughout the process. The data collection spanned about a month.

Both regions encountered unique challenges. In Amhara, farmer hesitation stemmed from fears of losing land to non-agricultural industries. There was also a prevailing distrust regarding the purpose of the collected data, given the significance of harvest piles for livelihoods.

Tigray presented a unique set of challenges. Many of the chosen sites had been active battlefronts in recent years, carrying high risks of unexploded bombs. Additionally, the team faced instances of dog attacks, particularly prevalent in the Central zone where dogs had not received vaccinations for approximately two years due to the conflict. Since the troops had not yet left

Tigray territory, the team faced exposure to troops from Amhara and Eritrea. There were also snake attacks in areas like the Menji-Guya line. The security situation was precarious and frightening during the field work.

Appendix Figure A5 illustrates the geographical distribution of the 2,296 data collection points acquired by our survey team. These points span across the Tigray and Amhara regions.

I. Closeup of ESA comparison

In Figure A6 we show close up examples of the locations in squares shown in Figure 9a, overlaying the ESA map in pink.

To accurately determine the additional cropland area projected by our model, we employed a systematic process. Surrounding each prediction point generated by our model, we established bounding boxes measuring 256x256 meters. Within these boxes, we evaluated the extent of coverage by the ESA cropmask, specifically targeting positive bounding boxes. If a given box had an ESA cropmask coverage of 20 percent or less, we classified it as newly predicted cropland by our model. For the shared cropland area recognized by both our model and ESA, we summed the areas of positive squares exhibiting an 80 percent or higher overlap with the ESA cropmask. Employing a similar methodology, we identified non-cropland areas mutually disregarded by both our model and ESA, by tallying the area of negative squares with an ESA cropmask coverage of 20 percent or lower.

Partition assignment code

Listing A1: Contiguous shape group partitioning algorithm

```
# Create a graph with rectangles as nodes and overlaps as edges
import pandas as pd
import os
import networkx as nx
from shapely.geometry import box
from shapely.strtree import STRtree

df = pd.read_csv(os.path.join(FOLDER_PATH, "merged_labelled.csv"))
df = df.iloc[:, 1:]

G = nx.Graph()

# Create shapes and nodes
def create_rectangle(row):
    return box(row[’lat_2’], row[’lon_1’], row[’lat_1’], row[’lon_2’])

geometry=[]
for index, row in df.iterrows():
    G.add_node(index)
    geometry.append(create_rectangle(row))

tree = STRtree(geometry)

# Add edges for each overlapping box
for idx, shape in enumerate(geometry):
    for intersecting in tree.query(shape):
        if not shape.touches(geometry[intersecting]) and idx != intersecting:
            G.add_edge(idx, intersecting)

connected_components = list(nx.connected_components(G))
groups_of_rectangles = [list(component) for component in connected_components]
```

Labelling procedure

We conducted a labeling procedure with the primary objective of optimizing accuracy and leveraging expert knowledge, while simultaneously expanding the scale of our labeled dataset. In Stage 1: knowledge distillation (Appendix Figure A4), we (coauthors) did a walkthrough of some examples guided by experts to familiarize ourselves with the appearance distribution of positive and negative examples of harvest piles. In Stage 2: high bandwidth labeling we focused on transferring a foundational proficiency to teach public labellers how to detect trivial examples of harvest activity. To achieve this, we instructed labellers by presenting multiple illustrations depicting harvest-related activities highlighted in red circles, of the same composition as shown in Appendix Figure A2. The illustrative samples were intentionally broad in classifying harvest piles; for instance, even strictly negative cases such as plastic tarps concealing sesame and accumulations of harvest remnants repurposed as animal feed were presented as affirmative instances of harvest piles. This inclusive approach was done to minimize false negative labels.

In Stage 2 we used public labellers to relabel 3792 negative examples that were previously labelled by coordinators but denoted by experts to have many false negatives. To promote dataset quality while minimizing costs, each image was presented to two labellers, who gave a binary label after reading the instructions. Details about the batch job are listed in Appendix Table A4. We chose to increase the quality of our workers by setting minimum requirements for their historic task approval rate and count. It is interesting to note that our entire batch job was completed within 4 hours and 45 minutes. The efficiency of MTurk's crowd-sourced labeling capacity open the prospects of automated quality control in significantly enhancing our labeling throughput.

Table A6: Labelling Job Details

Task details		Job completion status	
Reward per assignment	$0.01	Assignments completed	7584
Number of assignments per task 2	2	Average time per assignment	8 min 24 sec
Time allotted per assignment	1 hour	Creation time	June 30, 2023 9:56 AM PDT
Task expires in	2 days	Completion time	June 30, 2023 2:40 PM PDT
Auto-approve and pay workers in	3 days		
Worker Requirements		Cost summary	
Require workers to be masters	No	Total reward	$75.84
HIT approval rate %	Greater than 98	Fees to Mechanical Turk	$75.84
Number of HITs Approved	Greater than 50	Total cost	$151.68

By the end of the crowdsourced labelling step, we had 3792 SkySat images, each labelled by two labellers. For 437 of the images, the labellers both agreed the image did not contain piles. For 1708 of the images, the labellers agreed the image contained piles. For the remaining 1647 images where the labellers did not agree, we (the coauthors and project coordinators) manually labelled the images again, using our better knowledge of the appearance of harvest piles on SkySat images. After our manual pass through, we had 1997 positively labelled images and 1795 negatively labelled images.

The 1997 positively labelled images were then sent to Stage 3: Expert QA. Here, our subject experts manually reviewed each image that we decided were highly probable candidates for positive examples of harvest piles. After review, 341 of the 1997 images were labelled as positives, and the remaining were labelled as negatives. When we combined these updated labels with our dataset, we ended up with our current labelled dataset of 2547 positives and 2985 negatives.

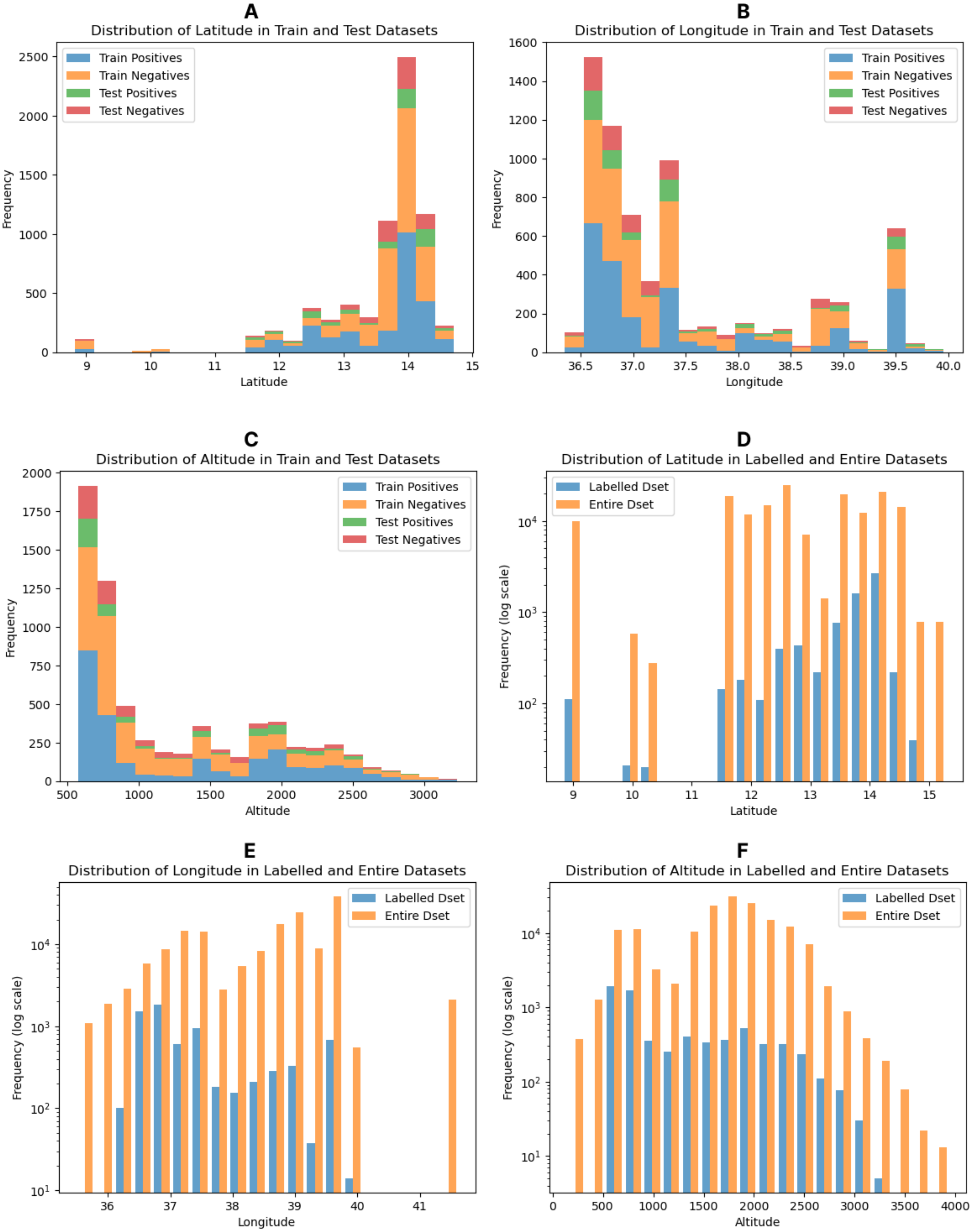


Figure A1: Statistics of HarvestNet dataset distribution

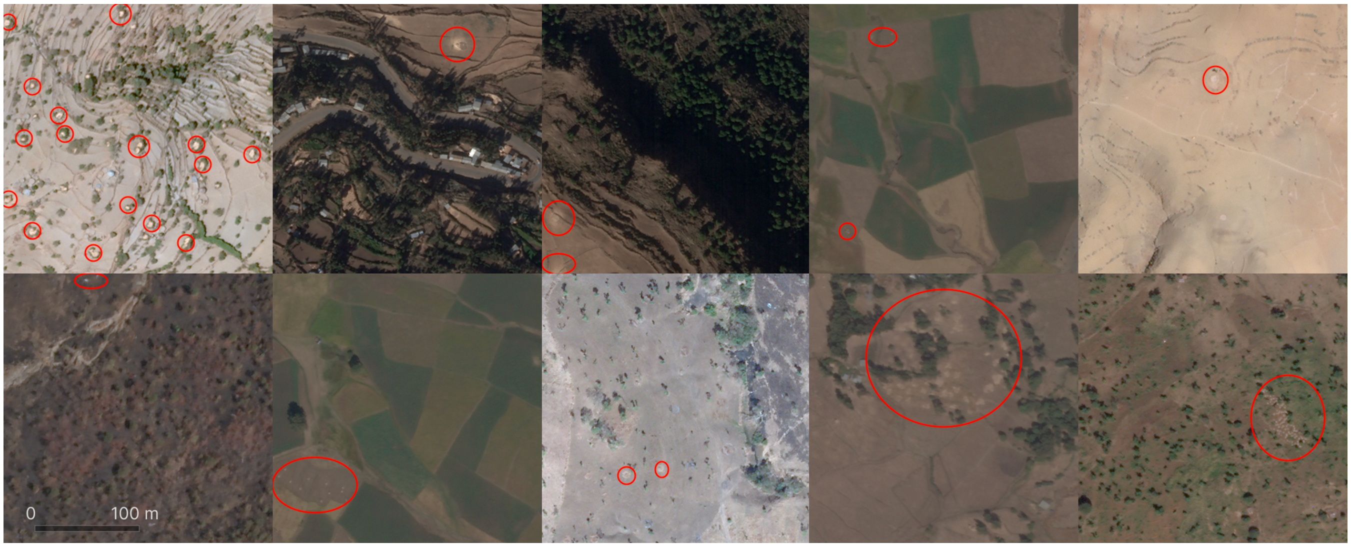


Figure A2: Examples of harvest piles at various stages, circled in red

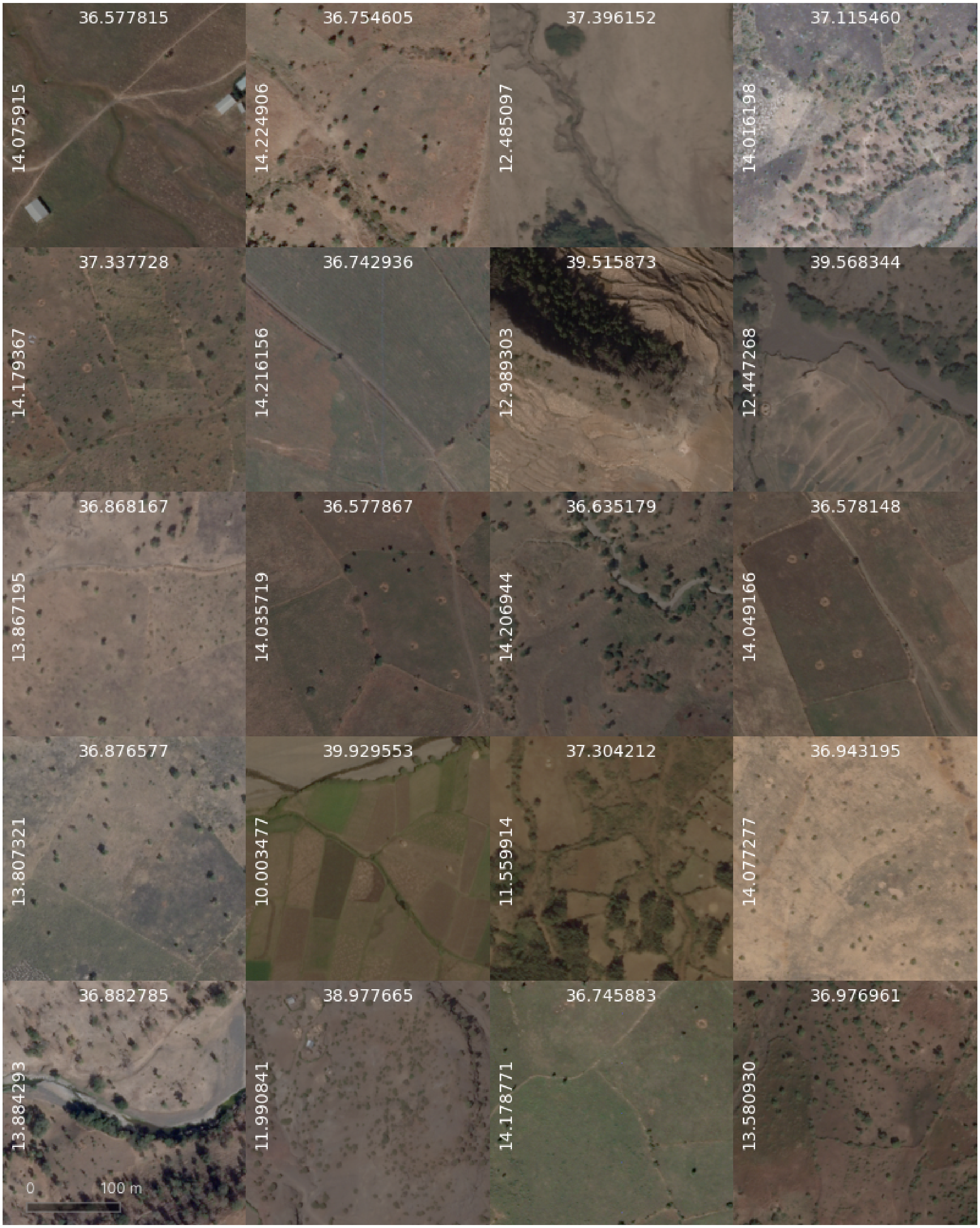
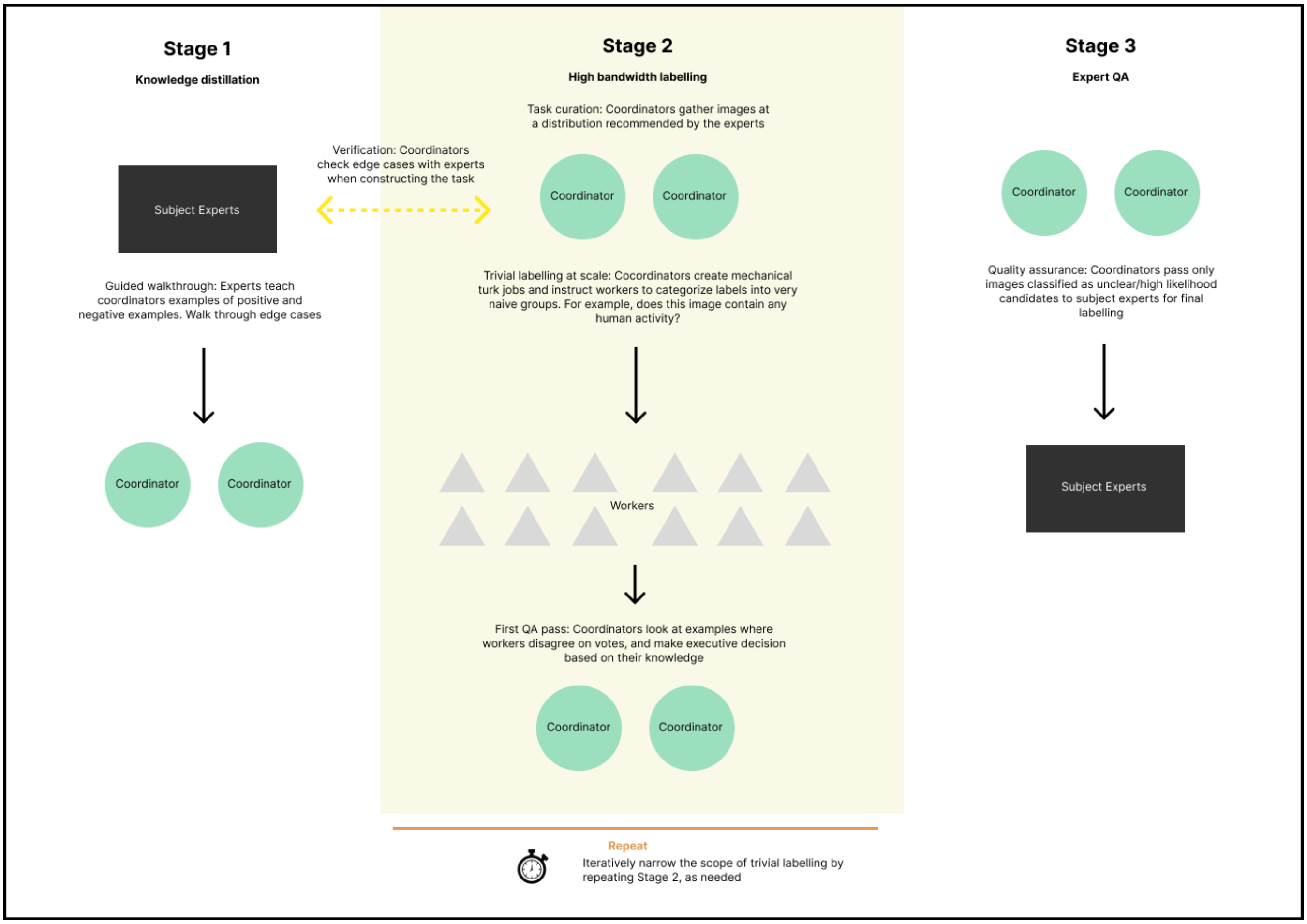


Figure A3: Additional examples of harvest pile activity, randomly selected



Guided walkthrough: Experts teach coordinators examples of positive and negative examples. Walk through edge cases

Verification: Coordinators check edge cases with experts when constructing the task

Task curation: Coordinators gather images at a distribution recommended by the experts

Trivial labelling at scale: Coordinators create mechanical turk jobs and instruct workers to categorize labels into very naive groups. For example, does this image contain any human activity?

First QA pass: Coordinators look at examples where workers disagree on votes, and make executive decision based on their knowledge

Quality assurance: Coordinators pass only images classified as unclear/high likelihood candidates to subject experts for final labelling

Iteratively narrow the scope of trivial labelling by repeating Stage 2, as needed

Figure A4: Labelling diagram

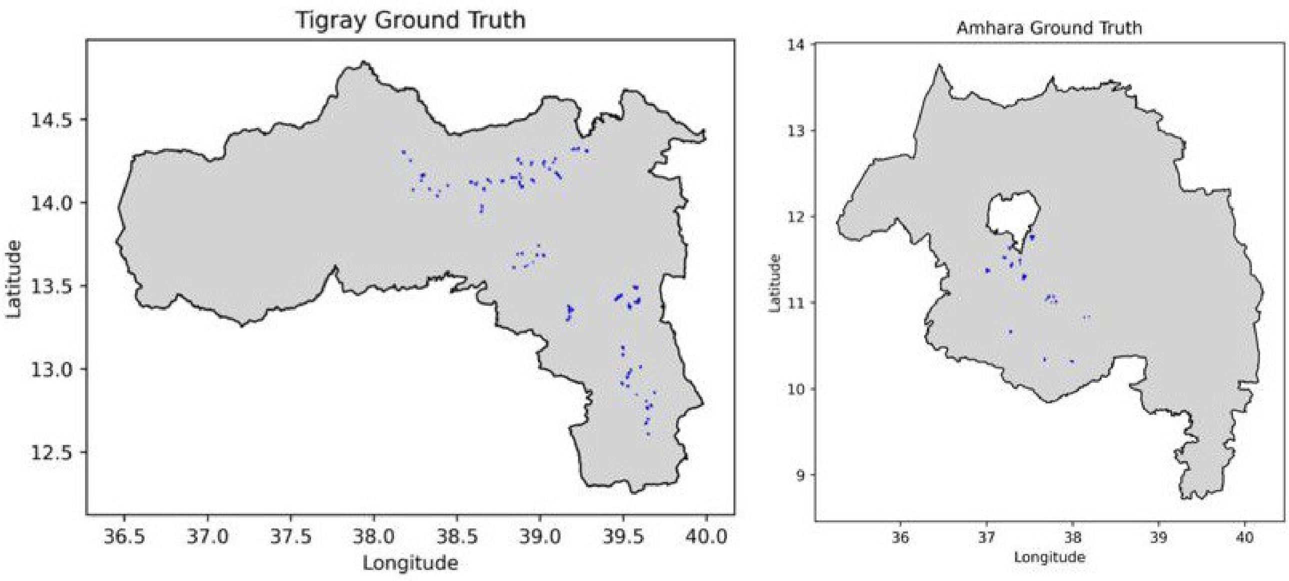
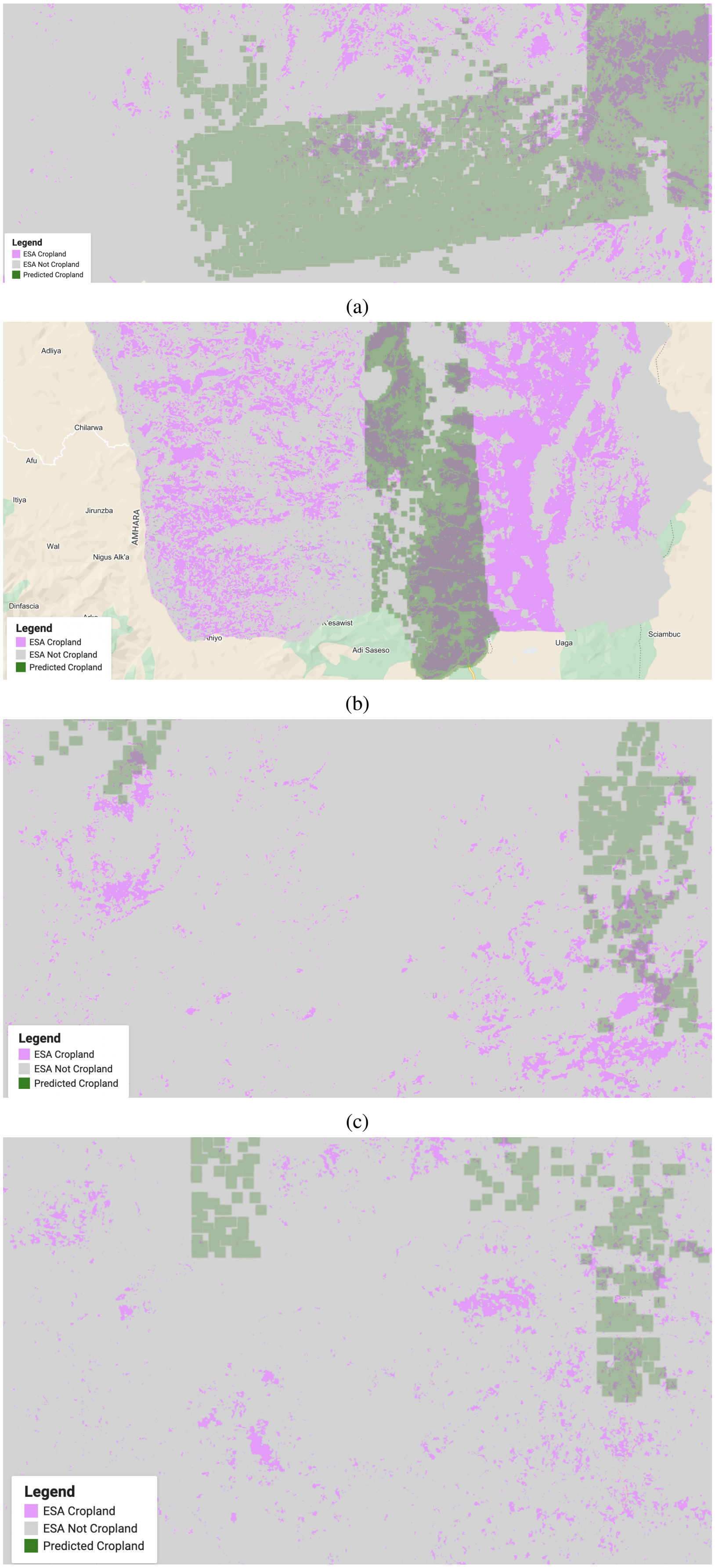


Figure A5: Ground truth collection zones in Tigray(left) and Amhara(right)



(a)

(b)

(c)

(d)

Figure A6: Close up view of ResNet-50 model predictions overlaid on top of ESA map