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

Supplementary Material

A. Appendix

A.1. 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.

A.2. 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://github.com/jonxuxu/harvest-piles.

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

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

A.3. 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.

A.4. Training parameters

In Appendix Table 1, 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.

A.5. Split counts

In Appendix Table 2, 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.

A.6. Ablation studies

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

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

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.

A.7. Dataset distribution

In Appendix Figure 1, 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 [1] had errors in lowlands in particular.

A.8. Examples of harvest piles

In Appendix Figure 2 and Appendix Figure 3 we provide more examples of harvest activity.

A.9. Ground truth collection

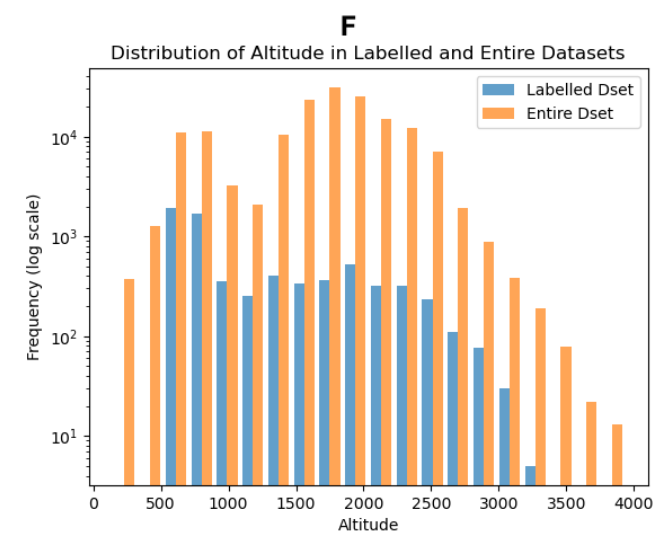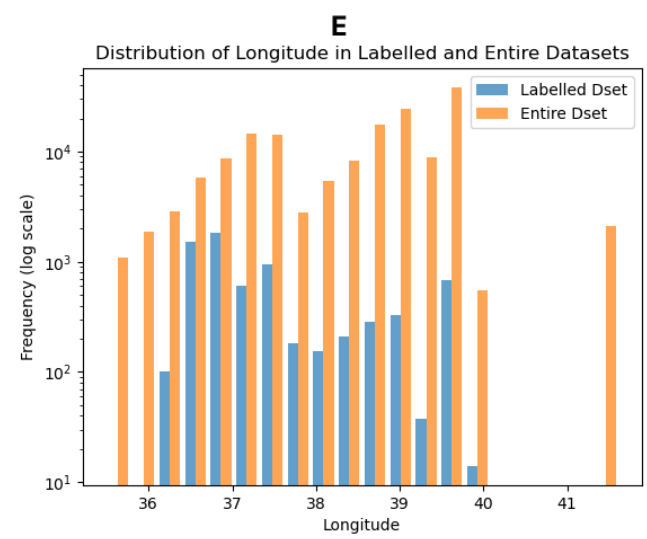During February and March 2023, we sent teams of six individuals to Tigray and Amhara regions respectively to collect
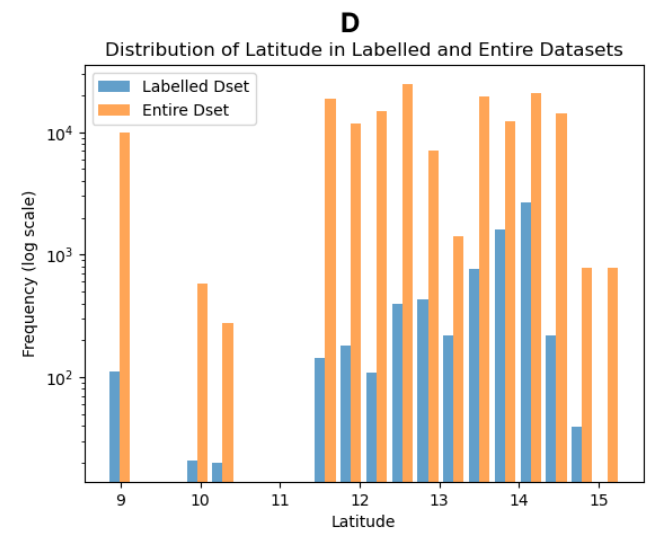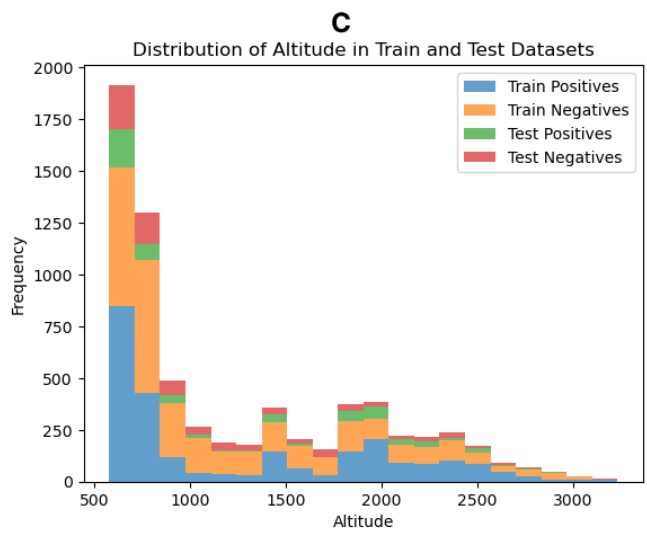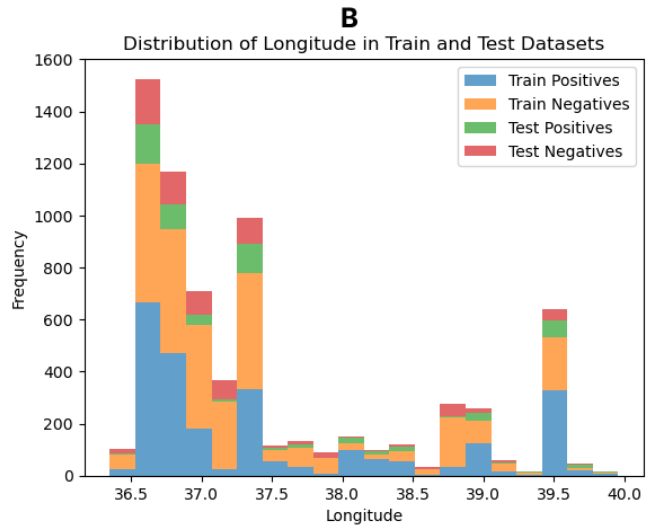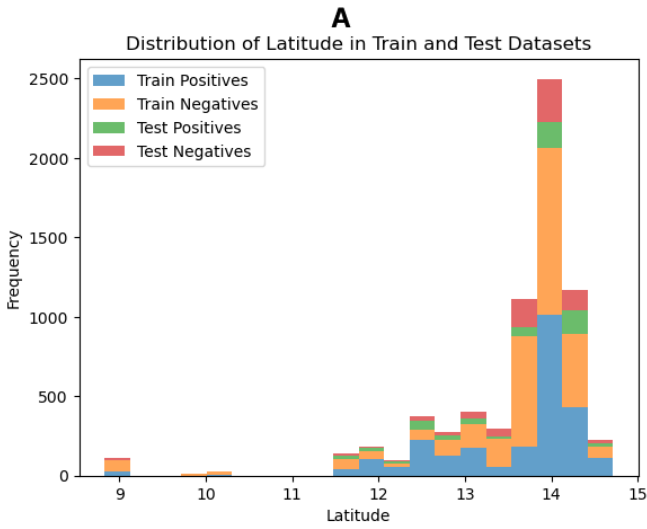
Figure 1. Statistics of HarvestNet dataset distribution

Table 1. 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

Table 2. 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

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 vaccina-

tions 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 5 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.

A.10. I. Closeup of ESA comparison

In Appendix Figure 6 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

Table 3. 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**

Table 4. 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**

Table 5. 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**

Figure 2. Examples of harvest piles at various stages, circled in red

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.

A.11. Partition assignment code

The code for assigning partitions is described in Appendix Listing 1.

A.12. 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 4), 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 2. 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 6. 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.

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.

Table 6. 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

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.

References

[1] Daniele Zanaga, Ruben Van De Kerchove, Dirk Daems, Wanda De Keersmaecker, Carsten Brockmann, Grit Kirches, Jan Wevers, Oliver Cartus, Maurizio Santoro, Steffen Fritz, et al. Esa worldcover 10 m 2021 v200. 2022. 1

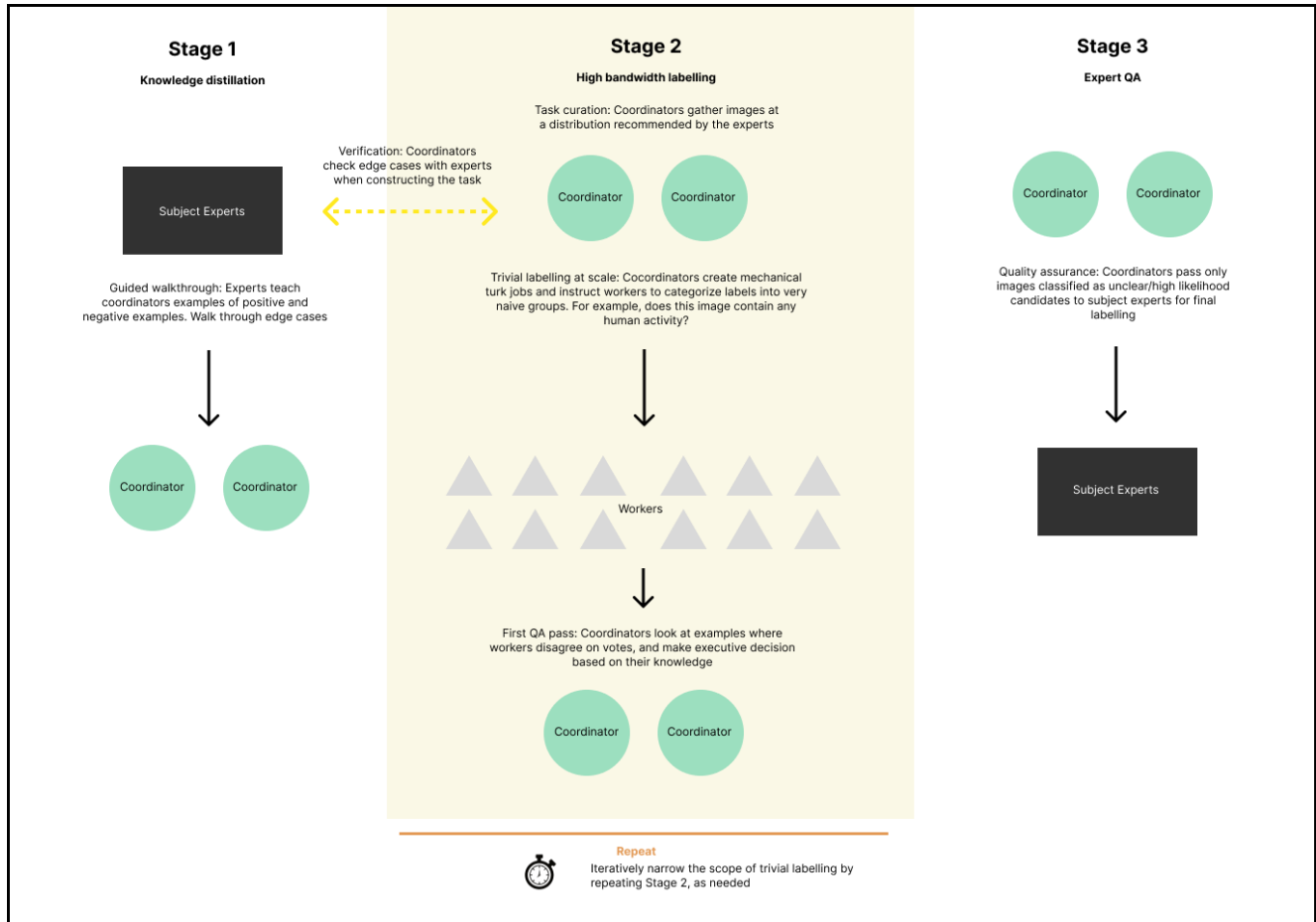Figure 3. Additional examples of harvest pile activity, randomly selected

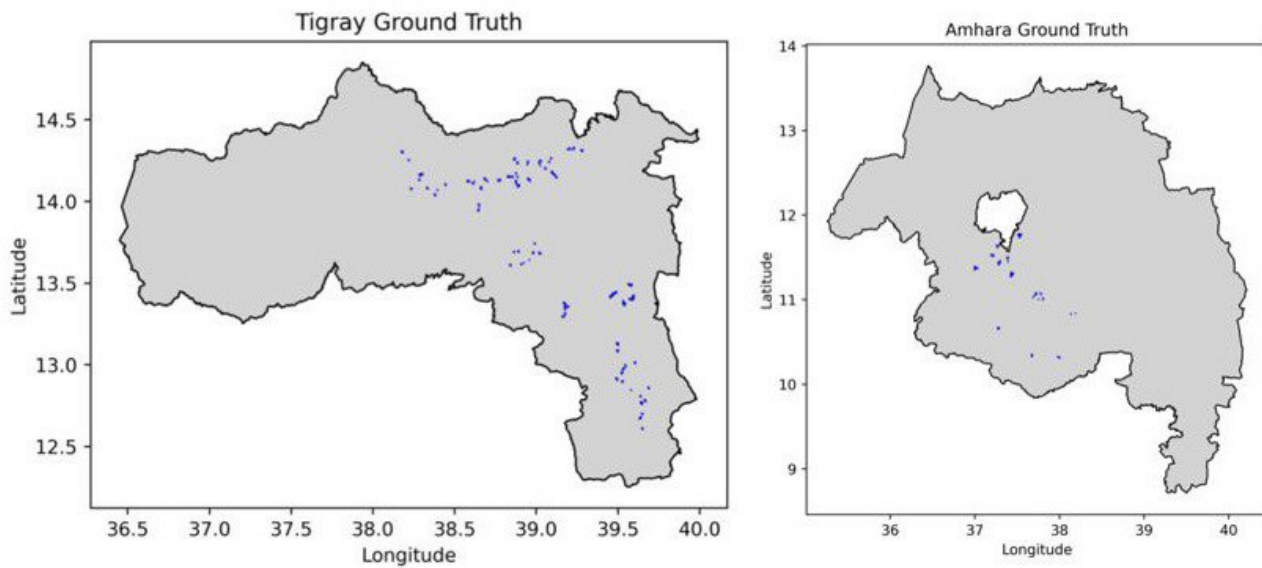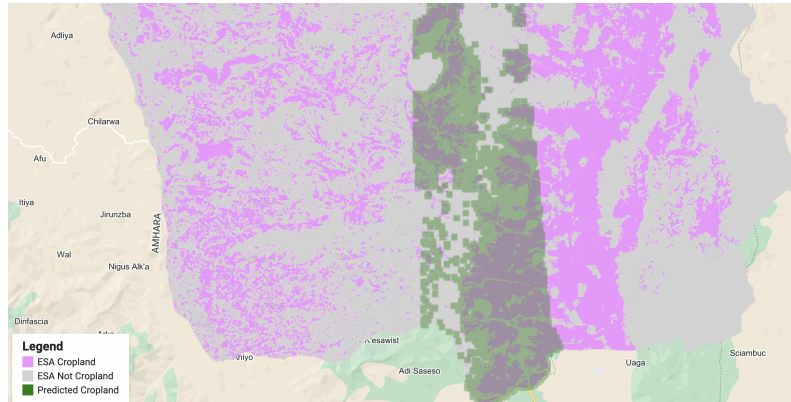Figure 4. Labelling diagram

Figure 5. Ground truth collection zones in Tigray(left) and Amhara(right)
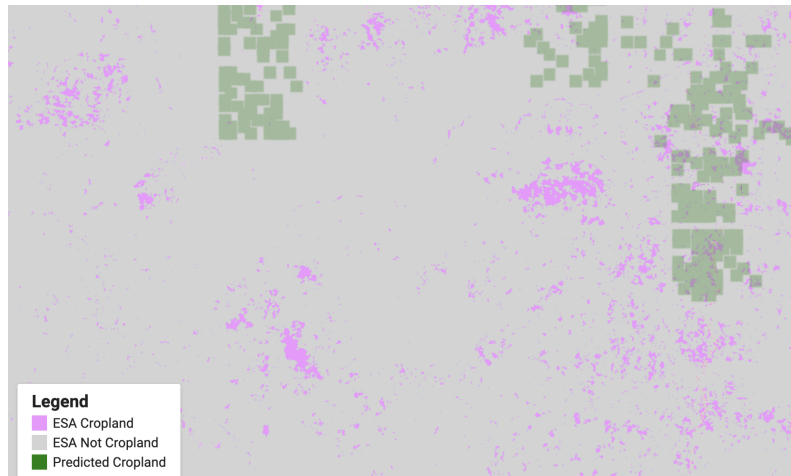
(a)

(b)

(c)

(d)

Figure 6. Close up view of ResNet-50 model predictions overlaid on top of ESA map

Listing 1. Contiguous shape group partitioning algorithm

```python
# 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]
```