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
j462xu@uwaterloo.ca, {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% and 98% accuracy on ground truth data for Tigray and 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. The dataset can be accessed through https://figshare.com/s/45a7b45556b90a9a11d2, while the code for the dataset and benchmarks is publicly available at https://github.com/jonxuxu/harvest-piles.

1. 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 [18, 24]. 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 [4, 5, 28] are inaccurate and updated infrequently in such regions, if at all.

Figure 1. Various examples of harvest piles.

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 [27]. Moreover, satellite images are now widely available at different resolutions with global coverage at low to no cost [20]. The performance of methods for mapping croplands in smallholder systems, however, remains limited in many cases [4, 28].

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 [8, 23]. In non-mechanized smallholder systems like Ethiopia, where subsistence rain-fed agriculture predominates [2], 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 sys-

*Equal contribution

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

Figure 3. Various stages of harvest activity.

tems, 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 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 [26]. 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 [28]. 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 [7, 25].

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-labeling 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 labeled 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 labeled 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 labeling 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 unlabeled data, and compared it against ESA WorldCover [28], one of the most updated land usage cover map according to [14].

2. Related Work

Mapping croplands using remote sensing has been well researched in the past [4, 5, 9, 12, 13, 16, 28]. Some methods use feature engineering with nonlinear classifiers [4, 12, 28], others use deep learning methods [13, 16]. 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 [28] and Dynamic World [4] 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 [1, 10, 13, 19] 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 [8, 23] have employed active learn-

ing 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.

3. 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 propose using RGB satellite imagery for pile detection due to its wide accessibility and adaptability for other uses.

3.1. Task Formulation

To demonstrate this method, we defined farmland detection as a binary classification task using square RGB satellite images at a set 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 consists 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 2 and 3.

3.2. 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.

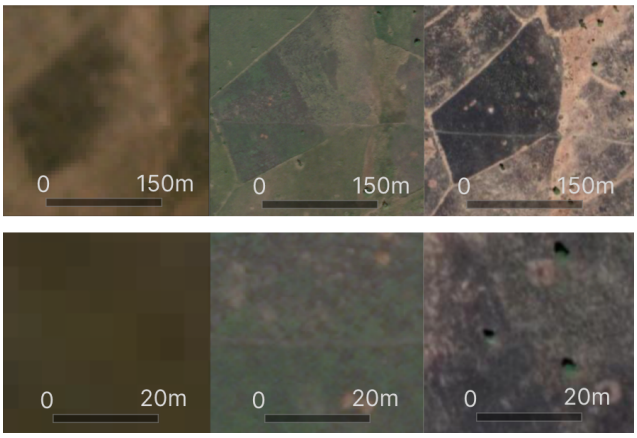

Figure 4. Side by side comparison of two areas, captured in 4.77m (left), 0.5m (center) and 0.3m (right) resolution. Note that piles become indistinguishable at 4.77m resolution.

3.2.1 Satellite Images

We use two image resolutions (0.5m and 4.77m per pixel) because the small size of harvest piles necessitates high-resolution images for accurate hand labeling and mapping, as Figure 4 shows.

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 unlabeled, 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. Each dataset entry includes unique latitude, longitude, altitude, and date, corresponding to SkySat images, with labeled examples also including PlanetScope images.

SkySat images [21] 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 [20] 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 high-

est 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-labeled 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.

3.2.2 Labeling

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. The majority of images from SkySat Collects contained no piles, so with the guidance of subject experts on agriculture in Tigray and Amhara, drew polygons outlining areas our experts knew had harvest piles (Figure 7). Downloading within those areas so, 30% of our scenes contained harvest piles.

The first stage of our labeling pipeline is to use crowdsourcing to filter out obvious negatives such as images consisting only of bare lands, and shrubs. Each image is shown to 2 anonymous Amazon Mechanical Turk workers (labeler details are described in Appendix Table 6, who are each tasked with deciding whether an image contains a pile. We teach the workers a very broad definition of a "pile" so that they filter out clear negatives without accidentally discarding potential positives. When one labeler votes no and another vote yes, we (the coordinators) cast the deciding vote. Afterwards, all images labeled as positive are forwarded to two experts (co-authors of the paper) for final evaluation by their consensus.

In Appendix Figure 4 we outline our labeling process in greater detail. The labeling process was done through inspection on SkySat images exclusively, afterwards PlanetScope images were paired with the corresponding labeled SkySat images. By the end of this stage, we had roughly 7k labeled 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.

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.

3.2.3 Ground Truth

In March 2023, we sent a survey team to collect ground truth data in Tigray and Amhara to validate our models' predictions 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 5. 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.

3.2.4 Dataset Split

Aiming for a balanced dataset, we targeted an 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 1.

To avoid contamination from overlapping images, we used graph traversal to form distinct groups. Each group consists of images that strictly overlap with at least one other image in the group. Images that did overlap any others

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

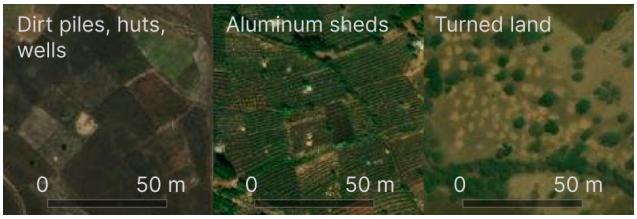

Figure 6. Examples of edge cases that are not harvest piles.

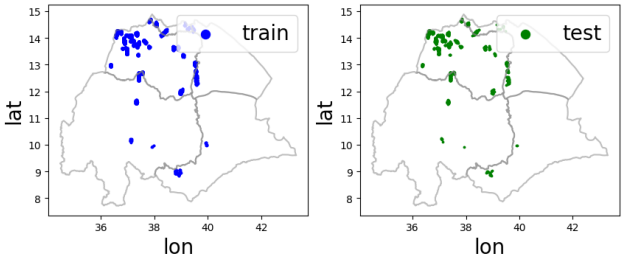

Figure 7. Training and test splits.

were assigned to their own individual groups. After graph traversal, our 6915 total images were divided into 5166 non-overlapping groups. Assigning each connected component a random color, the non-overlapping nature of groups are seen in Figure 8 A (the code is provided in Appendix Listing 1). Afterwards, we assigned each group between the train and test split. The largest group (776 images) and the second largest (323 images) were included in the training set, because they dominate a significant area in our dataset. Then, we randomly shuffled the remaining groups, and iteratively assigned them to either the train or test set to maintain a running 80:20 ratio between the train and test sets. This results in a train/test split (Figure 8 B) that does not overlap geographically, while still sharing a similar geographic distribution.

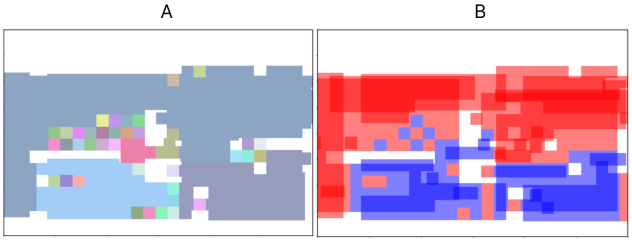

Figure 8. (A): An example region of image captures, organized into non-overlapping partitions of overlapping shapes, each assigned a random color. (B): Partitions are then divided into train (red) and test (blue).

3.3. Benchmarking

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

MOSAIKS [22] This approach uses non-deep learning to extract features from satellite images by convolving randomly chosen patches. These features are then utilized for downstream tasks, providing cost-effective performance. We featurize our dataset with 512 features per image and employ an XGBoost classifier for target prediction.

SATMAE [6] Based on masked autoencoders (MAE), this framework is pretrained on FMOW and Sentinel2 for various tasks, including single image, temporal, and multi-spectral. It exhibits strong performance in downstream and transfer learning tasks. We apply transfer learning by training this pre-trained model to predict the harvest pile's presence in our dataset.

Swin Autoencoder [17] is a vision transformer that creates hierarchical feature maps by merging patches in deeper layers and maintains linear complexity with image size via local window self-attention. It's pretrained with a masked image autoencoder on 150k Skysat images, scaling inputs to 224x224 pixels and partitioning them into 28x28 patches with a 40% mask ratio. A fully connected layer follows the transformer's 1x768 pooled output. The model is fine-tuned on our training set of labeled Skysat images.

Satlas [3] 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 [11] 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 labeled dataset.

4. Experiments

4.1. 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 1. 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.

4.2. 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 visually compare the ESA [28] land cover map with our ResNet-50-based classification map trained on HarvestNet. 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 within the area highlighted in Figure 9a. In Figure 9b, the positive classification (in green) of the best-performing model trained on our dataset overlays the ESA map (in pink), revealing our ability to detect new farmland in those regions. Figure 9d present satellite images of two example locations, verifying the presence of piles that were not detected by ESA and correctly identified in our map.

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.

4.3. Results

Table 1 shows our benchmark results from the HarvestNet dataset using hand-labeled test data. Table 2 presents the ResNet model's results on ground truth data. We use this model for its superior precision.

Figure 10 illustrates the Swin masked autoencoder's reconstruction, pretrained on 150k SkySat images. 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.

Figure 9 compares the ESA map (Figure 9a) with our predicted map (Figure 9b). We highlight specific areas, marked in black rectangles, where experts identified ESA's classification inaccuracies. A closer comparison of these regions of interest can be viewed in Appendix Figure 6. Satellite images from two locations (Figure 9d) show examples of ESA's inaccurate cropland detection. Table 3 lists the differences in our and ESA's map predictions. The aim is to compare

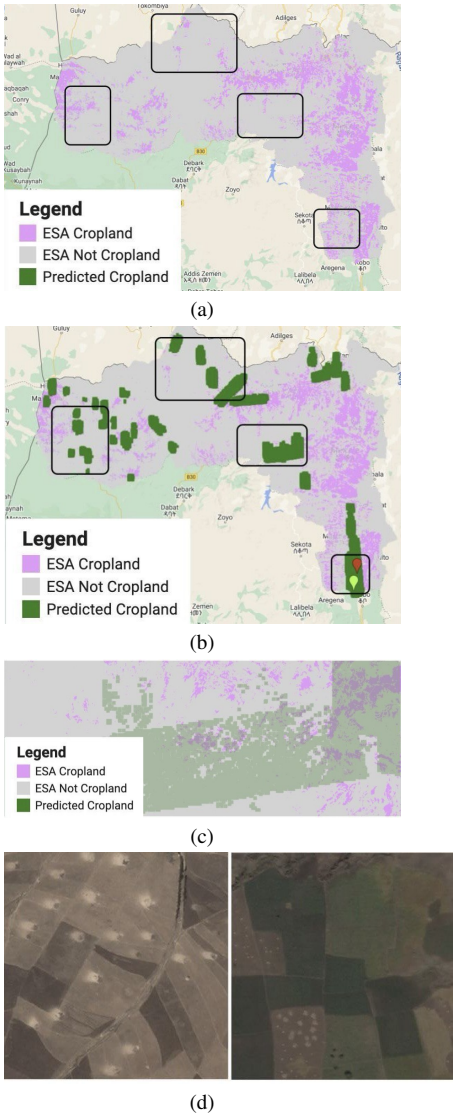

Figure 9. (a) The ESA map for our study region. (b) Positive predictions from our ResNet-50 model, overlaid the ESA map. (c) shows a zoom-in view of (b) in the southwest of Tigray region. (d) shows satellite images of the locations pinpointed in (b).

our number positive/negative predictions with those of ESA, while highlighting cases where our predictions differ. We also note the additional cropland area (in hectares) detected by our model and the overlapping cropland region shared between the two models.

5. 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

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.

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

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

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

	Total	- , +	+ , +	- , -
Samples	150,577	11,563	24,076	38,989
Area(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.

– 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 Table 2 reveals promising ResNet outcomes on ground truth data, attributed to two factors. Firstly, utilizing four distinct images per location, each representing a different month within the harvest season, enhances predictions by considering the union of results. This approach is feasible due to PlanetScope imagery's availability. Secondly, the ground truth data exclusively comprises positive samples, eliminating inherent false positives and contributing to elevated accuracy.

In Table 3 and Figure 9, we demonstrate our model's improvement over ESA predictions. Figure 9b highlights locations predicted as non-crop lands by ESA but as cropland by our model. Two samples of corresponding satellite images are shown in Figure 9d. Table 3 indicates instances where our model predicts farming activity and ESA does not (11k examples, 57k ha) and cases where both models predict similarly (62k samples, 199k ha). This added cropland is estimated by experts to be 90% true cropland, showcasing potential for improving existing maps in smallholder regions using harvest pile features. Appendix Figure 6 provides a higher zoom map for these missed locations.

Potential Bias Labelers may be biased in their interpretation of piles. Specific areas were chosen for image downloads based on pile presence, introducing sampling bias. Models trained on our dataset may exhibit bias toward processing piles with larger shapes and colors. Geographical bias exists in ground-collected data, chosen near roads for logistical reasons.

Limitations and Future Work Our study, focused on small feature classification due to resource limitations, utilized a dataset with certain constraints. We opted for binary labels over fixed 256x256 m areas for practical reasons, sacrificing spatial resolution for broader land coverage. Future improvements could include subdividing images for finer binary classification, leveraging negative sections as training data to enhance resolution.

Our approach was tailored for binary classification rather than detailed object detection or semantic segmentation, particularly of harvest piles. Adopting object detection could

provide valuable insights into pile locations, sizes, and densities, building on our initial goal to pinpoint areas with any pile presence. Exploring advanced models like Segment Anything [15] for automated segmentation presents a promising direction.

As seen in Figures 3, 5, and Appendix Figure 3, the study also identified various image features indicative of harvest activities as positive, which, with expert insights, could be further refined for improved object detection across diverse activity types.

Considering time series data could significantly enhance harvest pile detection. Despite our current focus on geographic diversity due to limited resources, integrating time-lapse imagery could improve model accuracy. HarvestNet's models, proven in this context, show promise for broader agricultural applications, such as detecting hay bales in North America, and may facilitate transfer learning across different agricultural domains.

6. 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.

References

[1] Siham Acharki. Planetscope contributions compared to sentinel-2, and landsat-8 for lulc mapping. *Remote Sensing Applications: Society and Environment*, 27:100774, 2022. 2

[2] Desale Kidane Asmamaw. 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, 2017. 1

[3] Favyen Bastani, Piper Wolters, Ritwik Gupta, Joe Ferdinando, and Aniruddha Kembhavi. Satlas: A large-scale, multi-task dataset for remote sensing image understanding. *arXiv preprint arXiv:2211.15660*, 2022. 5

[4] Christopher F Brown, Steven P Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, Valerie J Pasquarella, Robert Haertel, Simon Ilyushchenko, et al. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, 9(1):251, 2022. 1, 2

[5] Marcel Buchhorn, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, Luc Bertels, and Bruno Smets. Copernicus global land cover layers—collection 2. *Remote Sensing*, 12(6):1044, 2020. 1, 2

[6] Yezhen Cong, Samar Khanna, Chenlin Meng, Patrick Liu, Erik Rozi, Yutong He, Marshall Burke, David Lobell, and Stefano Ermon. Satmae: Pre-training transformers for temporal and multi-spectral satellite imagery. *Advances in Neural Information Processing Systems*, 35:197–211, 2022. 5

[7] ESS. 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, 2023. 2

[8] Lyndon D Estes, Su Ye, Lei Song, Boka Luo, J Ronald Eastman, Zhenhua Meng, Qi Zhang, Dennis McRitchie, Stephanie R Debats, Justus Muhando, et al. High resolution, annual maps of field boundaries for smallholder-dominated croplands at national scales. *Frontiers in artificial intelligence*, 4:744863, 2022. 1, 2

[9] Mark A Friedl, Douglas K McIver, John CF Hodges, Xiaoyang Y Zhang, D Muchoney, Alan H Strahler, Curtis E Woodcock, Sucharita Gopal, Annemarie Schneider, Amanda Cooper, et al. Global land cover mapping from modis: algorithms and early results. *Remote sensing of Environment*, 83 (1-2):287–302, 2002. 2

[10] Kwame Oppong Hackman, Peng Gong, and Jie Wang. New land-cover maps of ghana for 2015 using landsat 8 and three popular classifiers for biodiversity assessment. *International Journal of Remote Sensing*, 38(14):4008–4021, 2017. 2

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[12] Yulin Jiang, Zhou Lu, Shuo Li, Yongdeng Lei, Qingquan Chu, Xiaogang Yin, and Fu Chen. Large-scale and high-resolution crop mapping in china using sentinel-2 satellite imagery. *Agriculture*, 10(10):433, 2020. 2

[13] Hannah Kerner, Gabriel Tseng, Inbal Becker-Reshef, Catherine Nakalembe, Brian Barker, Blake Munshell, Madhava Paliyam, and Mehdi Hosseini. Rapid response crop maps in data sparse regions. *arXiv preprint arXiv:2006.16866*, 2020. 2

[14] Hannah Kerner, Catherine Nakalembe, Adam Yang, Ivan Zvonkov, Ryan McWeeny, Gabriel Tseng, and Inbal Becker-Reshef. How accurate are existing land cover maps for agriculture in sub-saharan africa? *arXiv preprint arXiv:2307.02575*, 2023. 2

[15] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023. 8

[16] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, 2017. 2

[17] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022. 5

[18] Sarah K Lowder, Jakob Skoet, and Terri Raney. The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World development*, 87:16–29, 2016. 1

[19] Sosdito Mananze, Isabel Pôças, and Mário Cunha. 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, 2020. 2

[20] Planet Labs. planet visual maps. https://developers.planet.com/docs/data/visual-basemaps/, 2023. Accessed: 2023-08-09. 1, 3

[21] planet labs. skysat docs. https://developers.planet.com/docs/data/skysat/, 2023. Accessed: 2023-08-09. 3

[22] Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1):4392, 2021. 5

[23] Philippe Rufin, Adia Bey, Michelle Picoli, and Patrick Meyfroidt. 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, 2022. 1, 2

[24] Leah H Samberg, James S Gerber, Navin Ramankutty, Mario Herrero, and Paul C West. Subnational distribution of average farm size and smallholder contributions to global food production. *Environmental Research Letters*, 11(12):124010, 2016. 1

[25] Teshome Sirany and Esubalew Tadele. Economics of sesame and its use dynamics in ethiopia. *The Scientific World Journal*, 2022, 2022. 2

[26] Statista. 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/, 2021. Accessed: 2023-08-09. 2

[27] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B Lobell, and Stefano Ermon. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*, 2021. 1

[28] 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, 2, 6