nasaharvest /
cropharvest

<> **Code**   ⊙ Issues 30   ⅄ Pull requests 4   ▷ Actions   ⊞ Projects   ⊙ Security   ⬃ Insights

Open source remote sensing dataset with benchmarks

⚖ CC-BY-SA-4.0 license

☆ **152** stars   ⅄ **37** forks   ⊙ **7** watching   ⅄ **11** Branches   ◇ **11** Tags   ⌁ Activity   ▤ Custom properties

⊕ Public repository

⅄   ⅄ **11** Branches   ◇ **11** Tags   ⅄   ◇   🔍 Go to file   t   Go to file   +   Add file ▾   Code   ···

👤 **gabrieltseng** Merge pull request #134 from nasaharvest/redeployment   •••   ✓

aa3c94a · 2 months ago   🕔

| 📁 .github/workflows	Run workflows on ubuntu-latest	2 months ago
📁 benchmarks	Fixes to script	7 months ago
📁 cropharvest	more mypy	2 months ago
📁 data	Add information about files to the R...	2 years ago
📁 diagrams	Add pipeline diagram to the readme	2 years ago
📁 process_labels	Apply along the right axis	8 months ago
📁 test	line length	2 months ago
📄 .flake8	pytest-flake8 keeps breaking - get ri...	2 years ago
📄 .gitignore	tanzania ecaas rice dataset fixed	2 years ago
📄 LICENSE.txt	Update README, add a contributin...	3 years ago
📄 MANIFEST.in	Add new line	3 years ago
📄 README.md	Update README instructions for Wi...	last year
📄 conftest.py	Add an column for the test datapoi...	3 years ago
📄 contributing.md	Reminder to update datasets.md in...	2 years ago
📄 datasets.md	Update datasets.md	9 months ago
📄 demo.ipynb	Notebook accessible on colab	2 years ago
📄 demo_exporting_data.ipynb	Exporting demo	2 years ago
📄 mypy.ini	Add an column for the test datapoi...	3 years ago

📄 pyproject.toml	Add toml new line	3 years ago
📄 release_steps.md	Remove unnecessary line	2 years ago
📄 requirements-benchmarks.txt	Remove learn2learn dependency	last year
📄 requirements-dev.txt	better pinning	2 months ago
📄 setup.py	New release	6 months ago
📄 zenodo_upload.sh	Zenodo upload script	2 years ago

📖 README   ⚖️ CC-BY-SA-4.0 license   ✏️ ☰

CropHarvest

CropHarvest is an open source remote sensing dataset for agriculture with benchmarks. It collects data from a variety of agricultural land use datasets and remote sensing products.
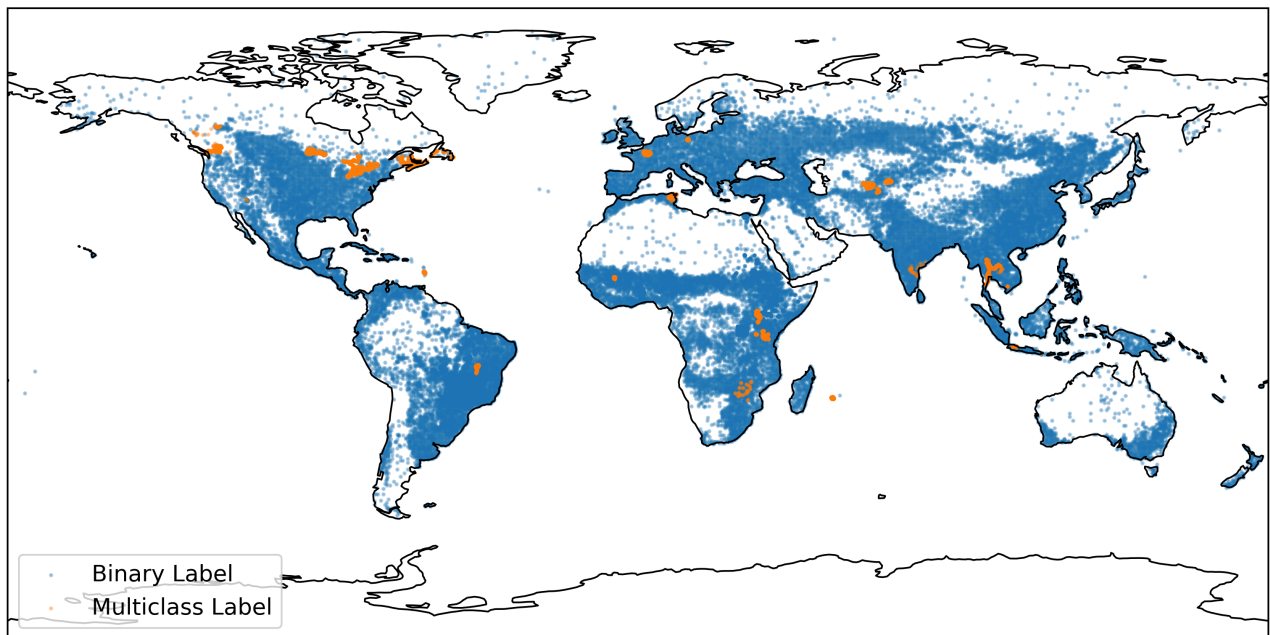
The dataset consists of **95,186** datapoints, of which **33,205** (35%) have multiclass labels. All other datapoints only have binary crop / non-crop labels.

70,213 (74%) of these labels are paired with remote sensing and climatology data, specifically [Sentinel-2](), [Sentinel-1](), the [SRTM Digital Elevation Model]() and [ERA 5 climatology data]().

21 datasets are aggregated into CropHarvest - these are documented [here]().

More details about CropHarvest and the benchmarks are available in [this paper]().

Pipeline

The code in this repository

1. combines the constituent datasets into a single geoJSON file,
2. exports the associated satellite data from Earth Engine,
3. combines both datasets to create `(X,y)` training tuples and
4. exposes those tuples via a `Dataset` object.

The pipeline through which this happens is shown below:
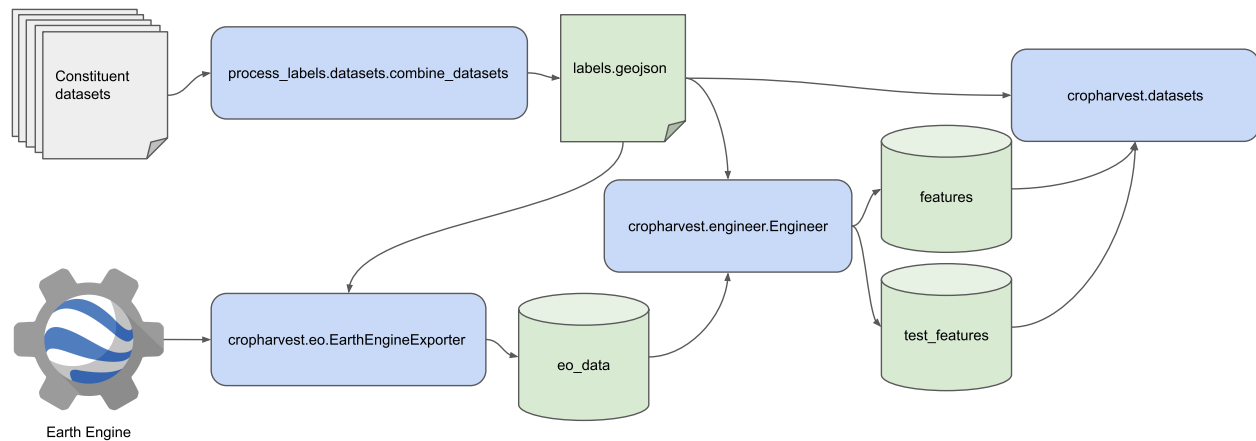
All blue boxes are associated with code in this repository. Anything green is data accessible via [Zenodo](#). By default, the Zenodo data will get downloaded to the [data folder](#) - the data folder's [Readme](#) has more information about the exact structure of the data.

There are unique cases where you may need to use the `EarthEngineExporter` directly, these use cases are demonstrated in the `demo_exporting_data.ipynb` notebook.

Installation

Linux and MacOS users can install the latest version of CropHarvest with the following command:

```
pip install cropharvest
```

Windows users must install the CropHarvest within a [conda](#) environment to ensure all dependencies are installed correctly:

```
conda install 'fiona>=1.5' 'rasterio>=1.2.6'
pip install cropharvest
```

In addition, it [may be necessary](#) to install `h5py` using conda as well (`conda install 'h5py>3.7.0'`) prior to pip-installing cropharvest.

Getting started CO Open in Colab

See the `demo.ipynb` notebook for an example on how to download the data from [Zenodo](#) and train a random forest against this data.

For more examples of models trained against this dataset, see the [benchmarks](#).

Contributing

If you would like to contribute a dataset, please see the [contributing readme](#).

~~FAQ~~ Questions asked at least once

▸ [How do I use CropHarvest for a specific geography?](#)

▸ [How do I load a specific pixel timeseries?](#)

▸ [What is the data format?](#)

License

CropHarvest has a [Creative Commons Attribution-ShareAlike 4.0 International](#) license.

Citation

If you use CropHarvest in your research, please use the following citation:

```
@inproceedings{
    tseng2021cropharvest,
    title={CropHarvest: A global dataset for crop-type classification},
    author={Gabriel Tseng and Ivan Zvonkov and Catherine Lilian Nakalembe and Hannah
Kerner},
    booktitle={Thirty-fifth Conference on Neural Information Processing Systems Datasets
and Benchmarks Track (Round 2)},
    year={2021},
    url={https://openreview.net/forum?id=JtjzUXPEaCu}
}
```

Releases 9

🏷 **v0.6.0** (Latest)
on Oct 31, 2022

+ 8 releases

Contributors 5

Languages

● **Jupyter Notebook** 75.9%   ● **Python** 24.0%   ● **Shell** 0.1%