Machine learning for harmful algal bloom detection
Harmful algal blooms (HABs) are a common threat to marine and human health.

Existing automated detection tools focus on ocean and coastal areas. But blooms in smaller inland water bodies are still monitored manually, which is very time intensive.

Image source: NASA Landsat Image Gallery
CyFi (Cyanobacteria Finder) is an open-source Python package that uses satellite imagery and machine learning to detect cyanobacteria levels, one type of HAB.

CyFi can help decision makers protect the public by flagging the highest-risk areas in lakes, reservoirs, and rivers quickly and easily.

*Image source: NASA Landsat Image Gallery*
CyFi uses high-resolution Sentinel-2 satellite imagery (10-30m) to focus on smaller water bodies with rapidly changing blooms.

Sentinel-3 is used by most existing tools, but its resolution of 300-500m is often too coarse for small, inland water bodies.
Generate estimates for many points at once with a simple CSV of dates and locations!

<table>
<thead>
<tr>
<th>date</th>
<th>latitude</th>
<th>longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 2023-06-27</td>
<td>41.287577</td>
<td>-80.424543</td>
</tr>
<tr>
<td>1 2023-07-10</td>
<td>35.650000</td>
<td>-78.682816</td>
</tr>
<tr>
<td>2 2023-08-31</td>
<td>35.705416</td>
<td>-79.164659</td>
</tr>
<tr>
<td>3 2023-09-22</td>
<td>37.564318</td>
<td>-101.335575</td>
</tr>
</tbody>
</table>

CyFi searches for and downloads publicly available satellite imagery around each point, which is passed into a machine learning model.
### Predicted cyanobacteria density (cells/ml)

<table>
<thead>
<tr>
<th>sample_id</th>
<th>date</th>
<th>latitude</th>
<th>longitude</th>
<th>density_cells_per_ml</th>
<th>severity</th>
</tr>
</thead>
<tbody>
<tr>
<td>89e12c14b5a131b82e9738932a7fa9c8</td>
<td>2023-06-27</td>
<td>41.287577</td>
<td>-80.424543</td>
<td>57,433</td>
<td>moderate</td>
</tr>
<tr>
<td>087d604d9d8568761513d26a47c94bc8</td>
<td>2023-07-10</td>
<td>35.650000</td>
<td>-78.682816</td>
<td>83,609</td>
<td>moderate</td>
</tr>
<tr>
<td>a0517780fa24874ebf166aefa17a0c1b</td>
<td>2023-08-31</td>
<td>35.705416</td>
<td>-79.164659</td>
<td>5,733</td>
<td>low</td>
</tr>
<tr>
<td>cde656c081bfe8fa99c7c8b20ff547f7</td>
<td>2023-09-22</td>
<td>37.564318</td>
<td>-101.335575</td>
<td>3,684,003</td>
<td>high</td>
</tr>
</tbody>
</table>

Cyanobacteria estimates are saved out as a CSV that can be plugged into any existing decision-making process.

For each point, the model provides a severity level based on World Health Organization (WHO) guidelines and an estimated density in cells per mL for detailed analysis.
Simply run one line of code to generate predictions

$ cyfi predict list_of_points.csv

SUCCESS | Loaded 5 sample points (unique combinations of date, latitude, and longitude) for prediction
SUCCESS | Downloaded satellite imagery
SUCCESS | Cyanobacteria estimates for 4 sample points saved to preds.csv
Or estimate cyanobacteria for a single point rather than providing a file.

```
$ cyfi predict-point --lat 35.6 --lon -78.7 --date 2023-09-25

SUCCESS | Estimate generated:
date       2023-09-25
latitude   35.6
longitude  -78.7
density_cells_per_ml  22,836
severity   moderate
```
Launch the CyFi Explorer to view cyanobacteria estimates alongside Sentinel-2 imagery!
CyFi was trained and evaluated using in-situ measurements of cyanobacteria density from across the U.S.

Train set: 8,979 observations
Test set: 4,035 observations
CyFi is most accurate at low and high cyanobacteria densities

**Low severity**
Better allocate ground sampling resources by deprioritizing water bodies where blooms are likely absent

**High severity**
Support public health interventions by flagging water bodies where severe blooms are likely present
A true positive (bloom presence) is where cyanobacteria density > 10,000 cells/mL. Uses a dataset of 756 ground measurement observations from across the U.S.

CyFi performs at least as well as Sentinel-3 based tools and has 10x greater coverage of lakes across the U.S. thanks to Sentinel-2 imagery!

Bloom detection accuracy

Water bodies detected by Sentinel-2 across the U.S. Source: Global Water Bodies Product
CyFi makes it simple for water quality managers to take advantage of state-of-the-art machine learning.

Plus, the algorithm is open source so anyone can reuse, update, or contribute.
To learn more and start using CyFi today, go to:

cyfi.drivendata.org