

The background image is an aerial photograph of a coastal region. On the left, a dense urban area with numerous buildings and roads is visible. The coastline curves from the bottom left towards the center. The water is a deep blue-green color, with large, swirling patches of bright green algae bloom. Some small islands or rocky outcrops are scattered in the lower right portion of the image.

Safeguarding Against Algae Bloom Threats

AquaWatch

Sally Fang, Yin Yin Teo, Lynda Solis
Chavez, Jonathan Phan

Our Team



Sally Fang

sallyfang@ischool.berkeley.edu

ML Lead

Model Evaluation Lead

Data Engineer / MLOps backup



Yin Yin (Jasmine) Teo

teoyinyin@ischool.berkeley.edu

Project/Product Manager

Chief Facilitator

ML Lead backup



Lynda Solis Chavez

lsolischavez@ischool.berkeley.edu

EDA Lead

Resident SME

Product / Project Manager backup



Jonathan Phan

jonathanphan@ischool.berkeley.edu

Data Engineer/MLOps

Lead MVP application developer

EDA lead backup

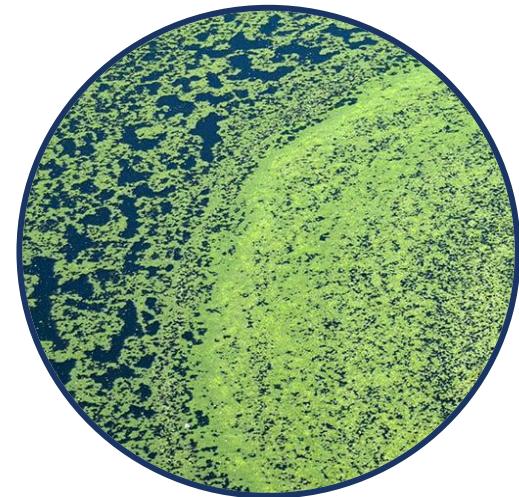
Understanding the Algae Problem



Irritation & Illness



Pets, livestock, & wildlife



Habitats, water quality,
& recreation

Microcystin Levels of Algae Bloom

No Advisory



< 0.8 $\mu\text{g}/\text{L}$

Caution



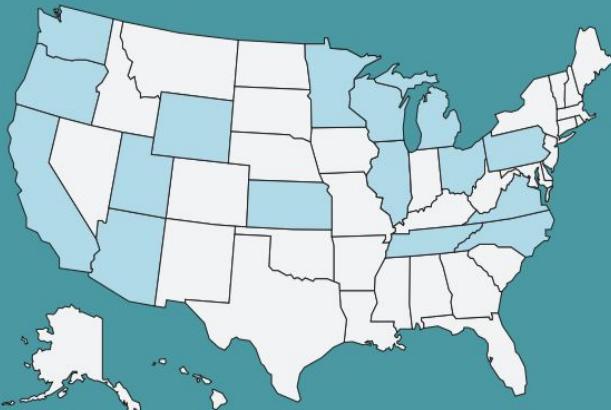
0.8 $\mu\text{g}/\text{L}$ to 20 $\mu\text{g}/\text{L}$

Danger



$\geq 20\mu\text{g}/\text{L}$

Harmful Algae Bloom (HAB) Crisis



16 states reported data for 2021

368 Harmful Algal Bloom Events



117 Human Illnesses



2,715 Animal Illnesses

Harmful Algal Bloom Events



9 out of 10
HAB events were in
fresh water, like lakes



HAB events
peaked in **August**

Most (85%)
were classified
as confirmed events²



92%
of ill
animals
died



About half of
ill people were **under**
the age of 18 years



74% sought medical care

Images and Statistics acquired from [cdc.gov](https://www.cdc.gov)

Market Opportunity

Primary Stakeholders:

1. Environmental Organizations (EPA, NOAA, USGS)
2. Health Organizations (CDC, NIH)
3. General Public

PR Newswire reports that the algaecides market in 2021 was US\$2.364 billion and is expected to grow to a size of US\$3.016 billion in 2028. This is a compound annual growth rate of 3.54%.



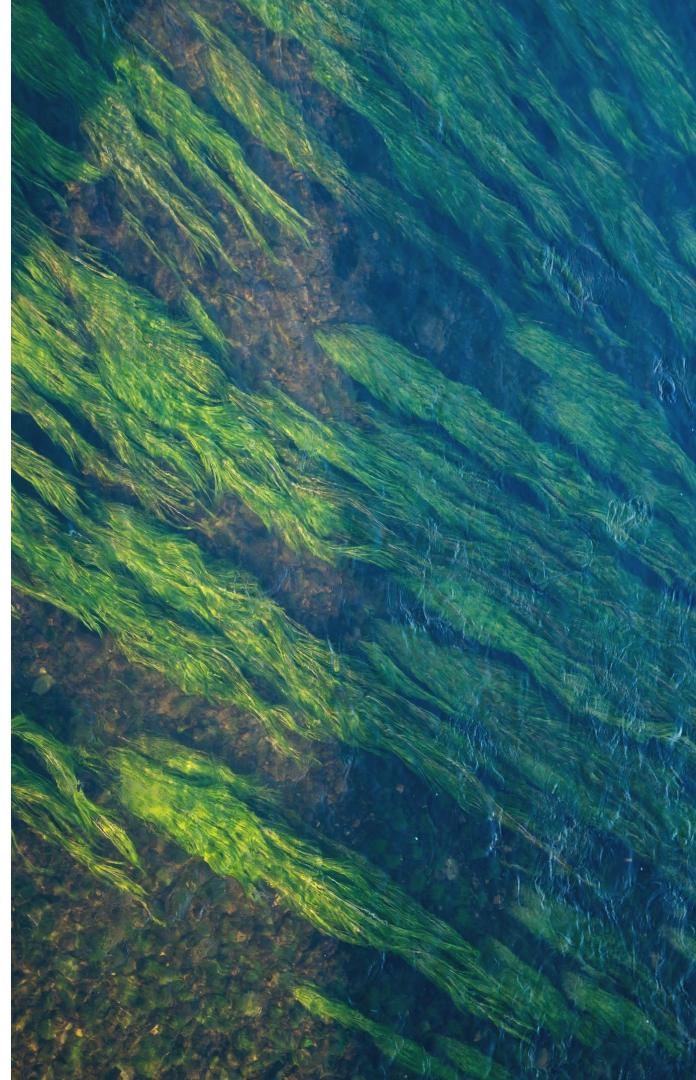
Our Solution

Develop an advanced image classification model
to **identify harmful algae** and **predict toxicity levels** in water bodies.

Impact

Early detection and timely responses to harmful algae blooms for ecosystem conservation.

Prediction of toxicity levels for **safeguarding public health** and **issuing warnings**.



Minimum Viable Product (MVP)

Key Questions:

1. How can we **quickly identify** harmful algae in water bodies?
2. How does the app address potential **image distortions** that may affect the accuracy of algae detection in **user-uploaded** photos?
3. Is it possible for **experts** to upload their records? What measures are in place to ensure the **reliability** of these records?

Reliable
Tool

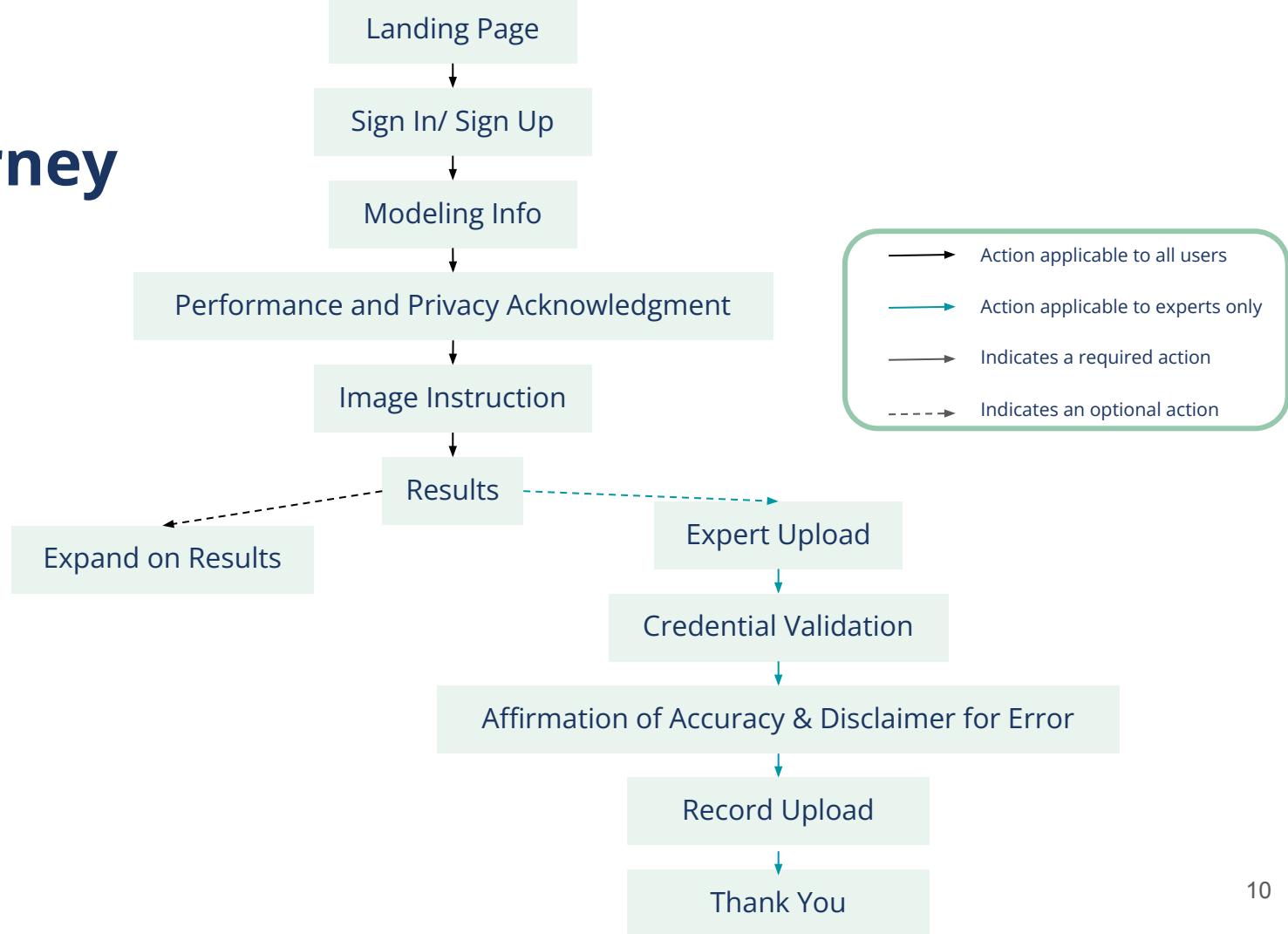


Quick
Detection



Informed
Decisions

MVP User Journey



- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- -> Indicates an optional action

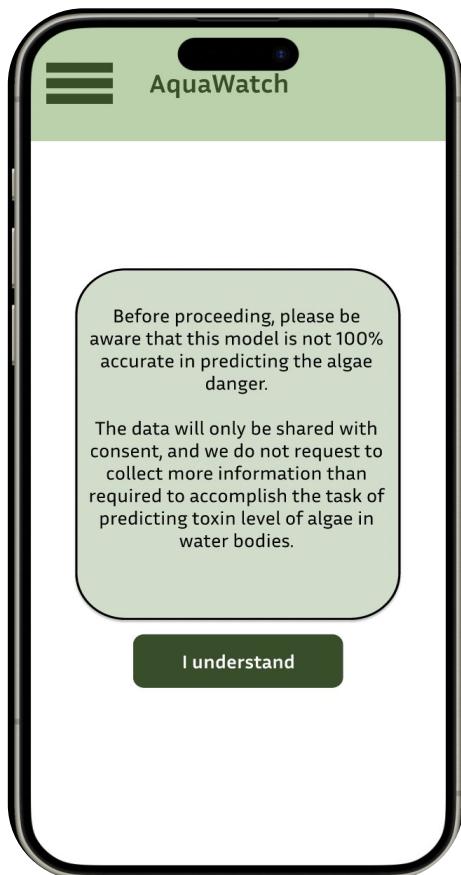
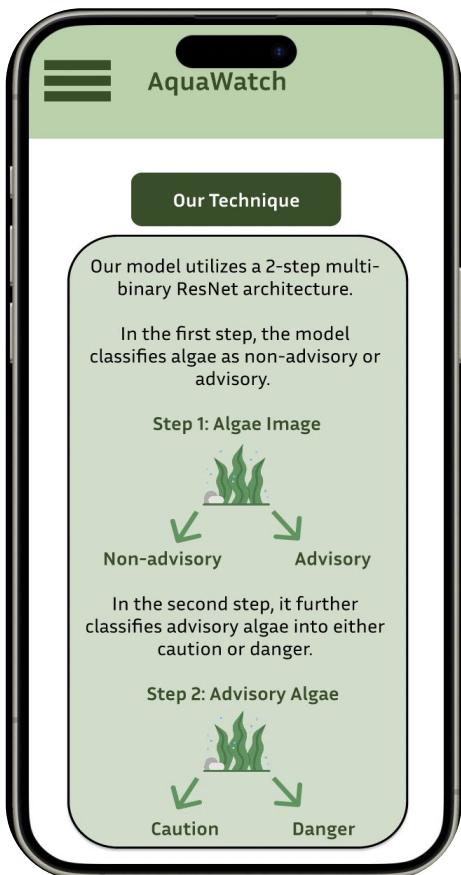
Landing Page

Sign In/ Sign Up

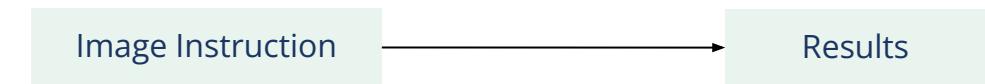


Modeling Info

Performance and Privacy Acknowledgment



- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- > Indicates an optional action



- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- > Indicates an optional action

Results

Expand on Results

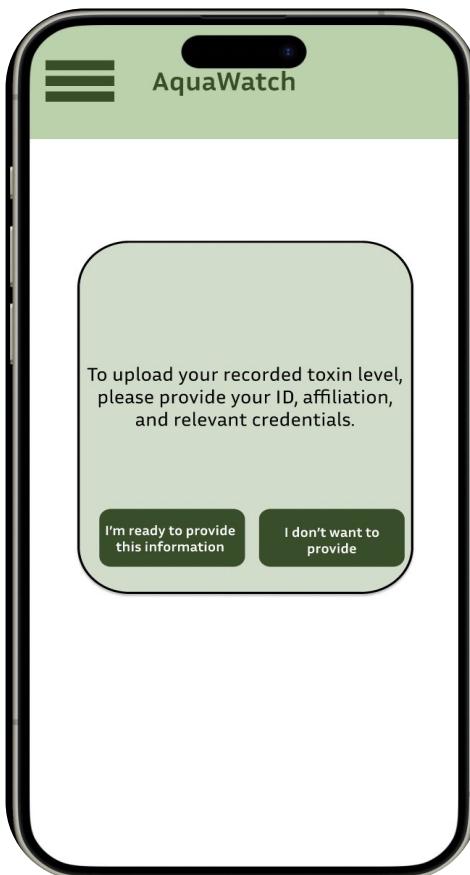
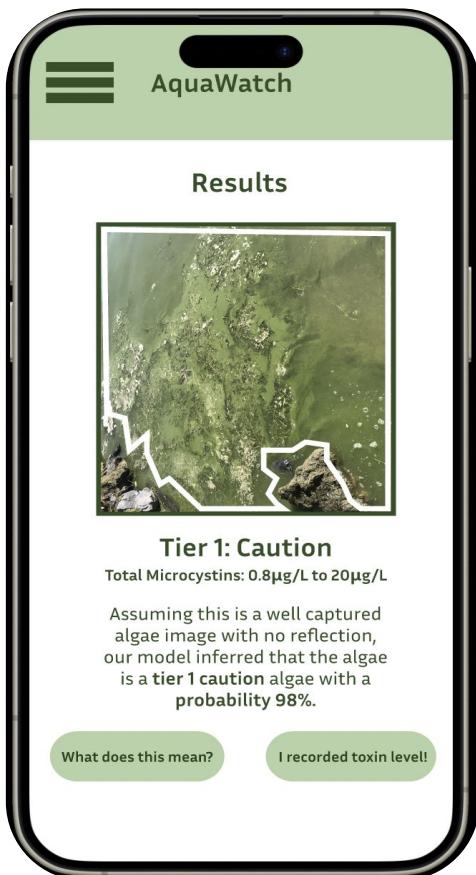
- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- > Indicates an optional action



Results

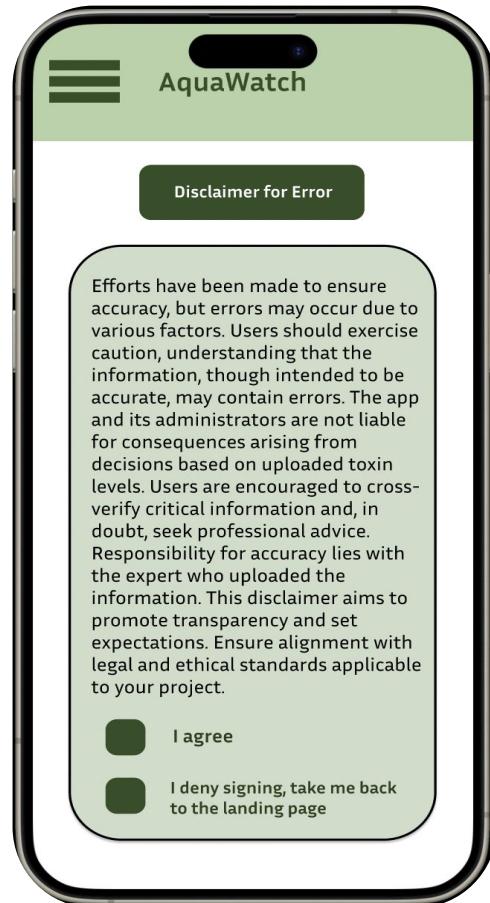
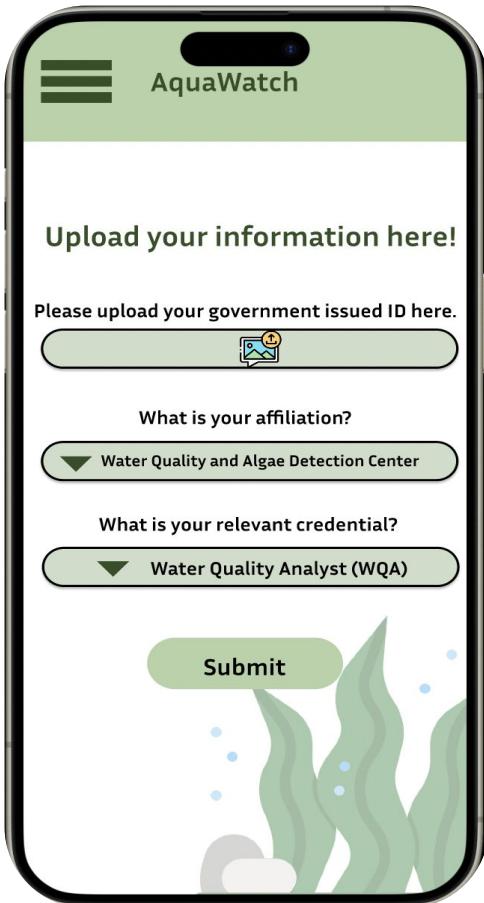
Expert Upload

- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- > Indicates an optional action



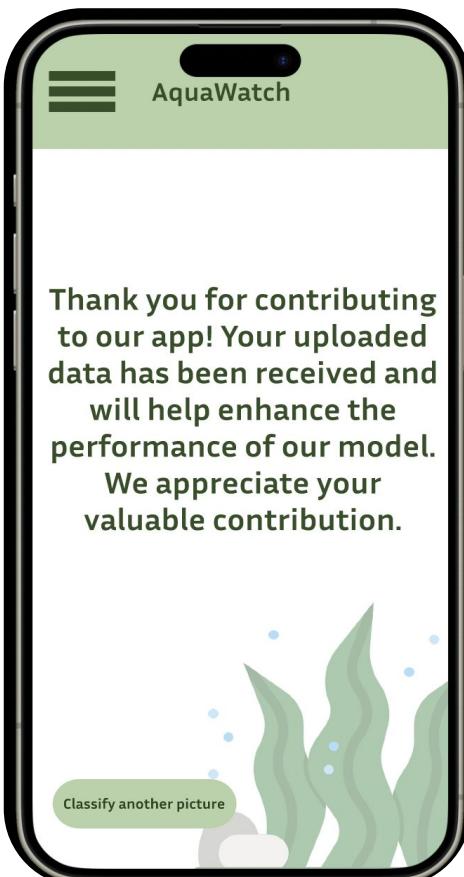
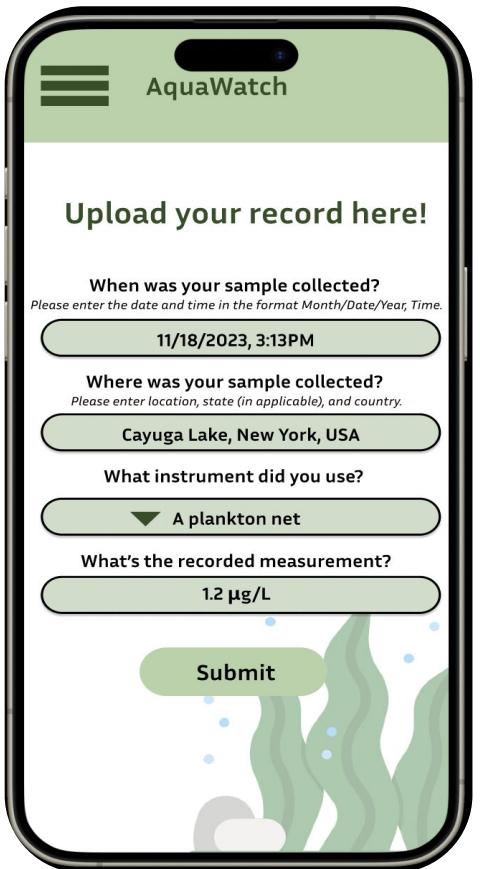
Credential Validation

Affirmation of Accuracy & Disclaimer for Error



Record Upload

Thank You



- Action applicable to all users
- Action applicable to experts only
- Indicates a required action
- > Indicates an optional action

MVP - Demo

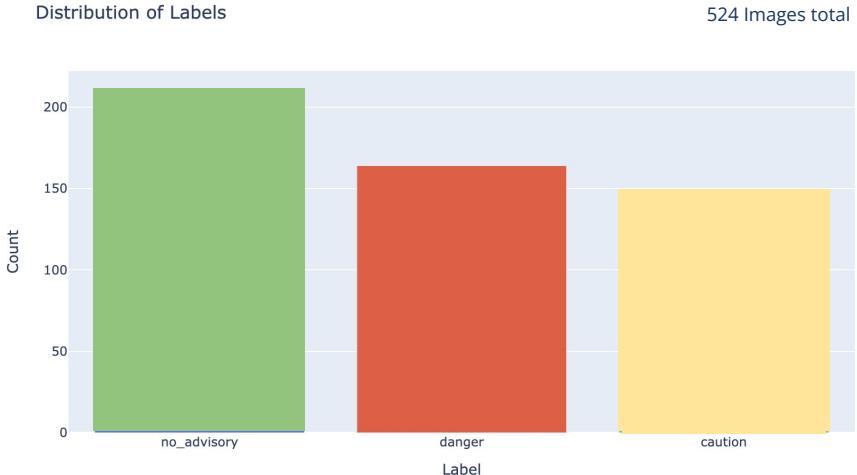


Two Scenarios

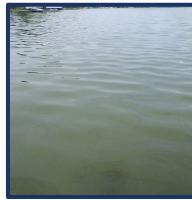
1. I bring my dog to Cayuga Lake and notice algae in the water. Let's see how I use the app to determine if it's safe for us to play in the water.
2. As an expert, I collect algae samples to record toxin levels. Let's see how I use the app to compare the results.

The Data

- Harmful Algal Bloom Datasets
- Clear Lake Cyanotoxin Issues
- Harmful Cyanobacterial Bloom (HCB) Advisories in Wyoming Waters
- Algal Bloom Sampling Status in Florida Water



No Advisory: < 0.8 μ g/L



Caution: 0.8 μ g/L to 20 μ g/L



Danger: $\geq 20\mu$ g/L



Approach



Imbalanced Class Distribution

Image Sizes

Multiclass vs Binary Model

Resize images to (224,224)

Convert image to pixels

Normalize pixel values

Data Augmentation

- Rotating, shifting, and flipping
- Adjusting brightness and contrast

Histogram of Oriented Gradients (HOG)

Local Binary Pattern (LBP)

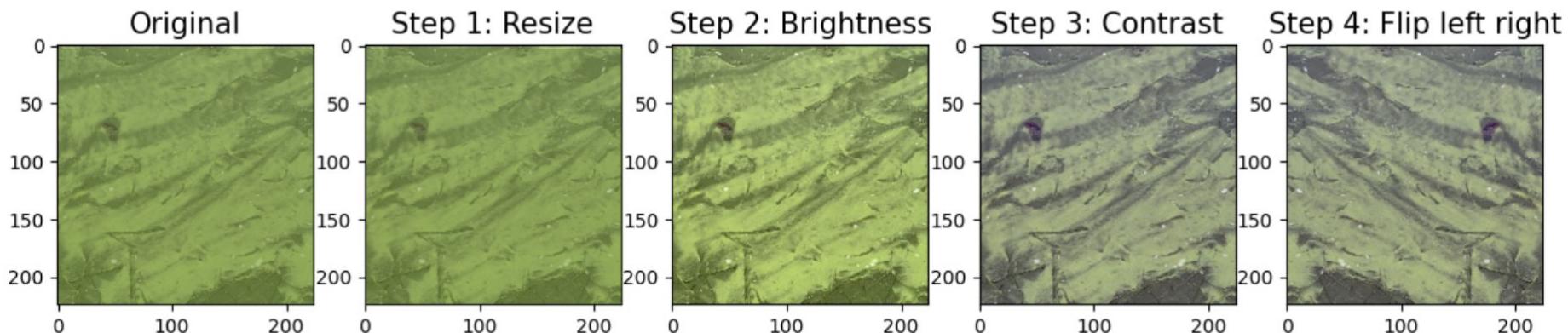
Stacked CNN Model

- No Advisory vs. Caution & Danger
- Caution vs. Danger

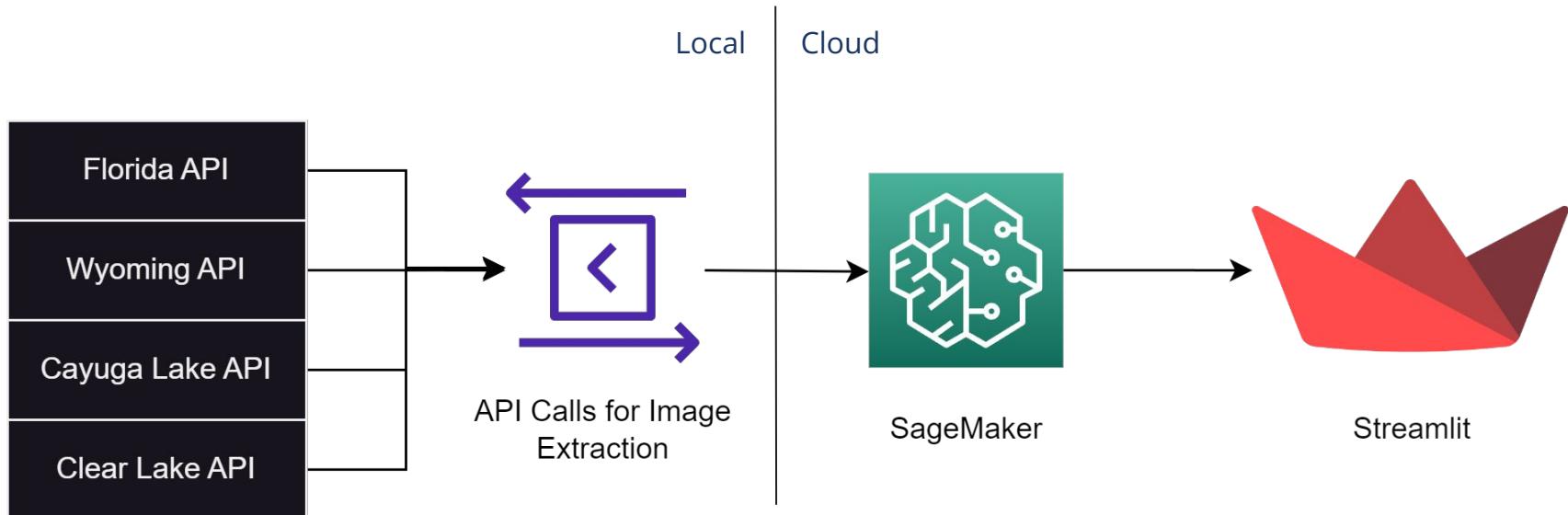
Stacked ResNet Model

- No Advisory vs. Caution & Danger
- Caution vs. Danger

Data Augmentation

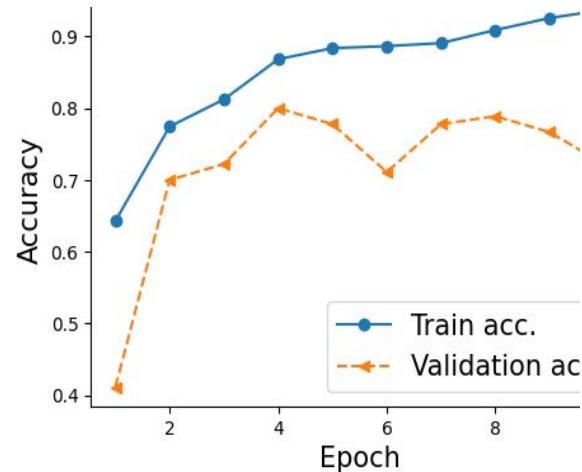
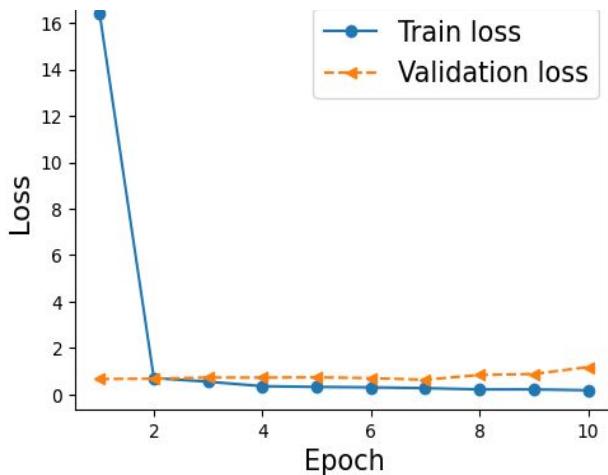


Data Pipeline



Baseline – Binary CNN

- 0 - No advisory (0.8ug < microcystins)
1 - Some algae levels



Image, 96x224x224

Conv2D, 5x5, 32, ReLU

MaxPooling2D,2x2

Flatten

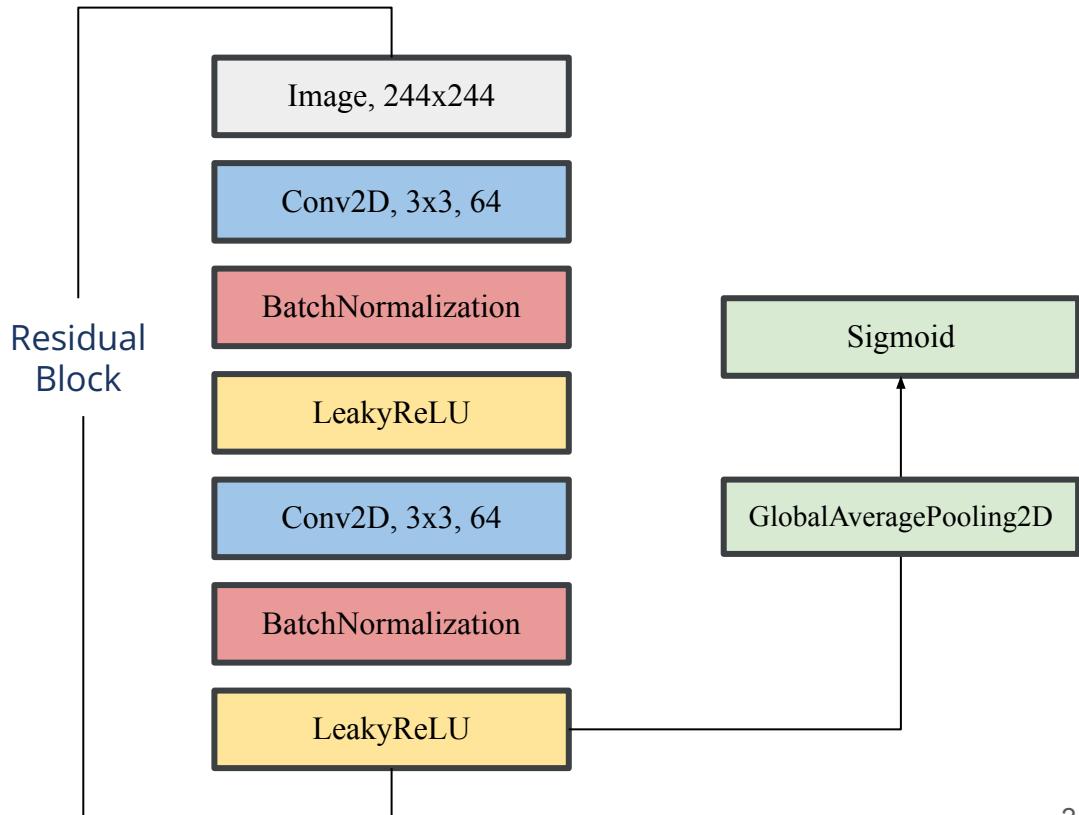
Dense, 1028, ReLU

Dropout Layer

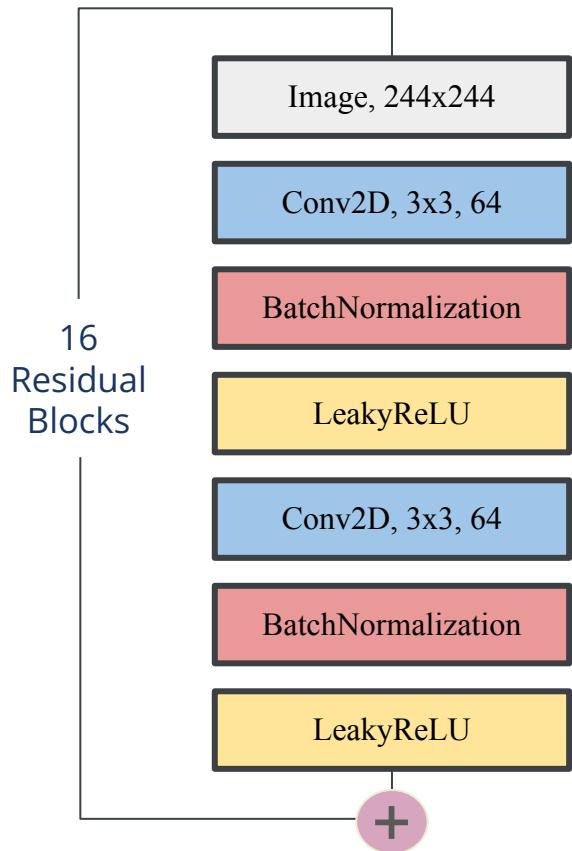
Dense, 1, None

Resnet Model Implementation

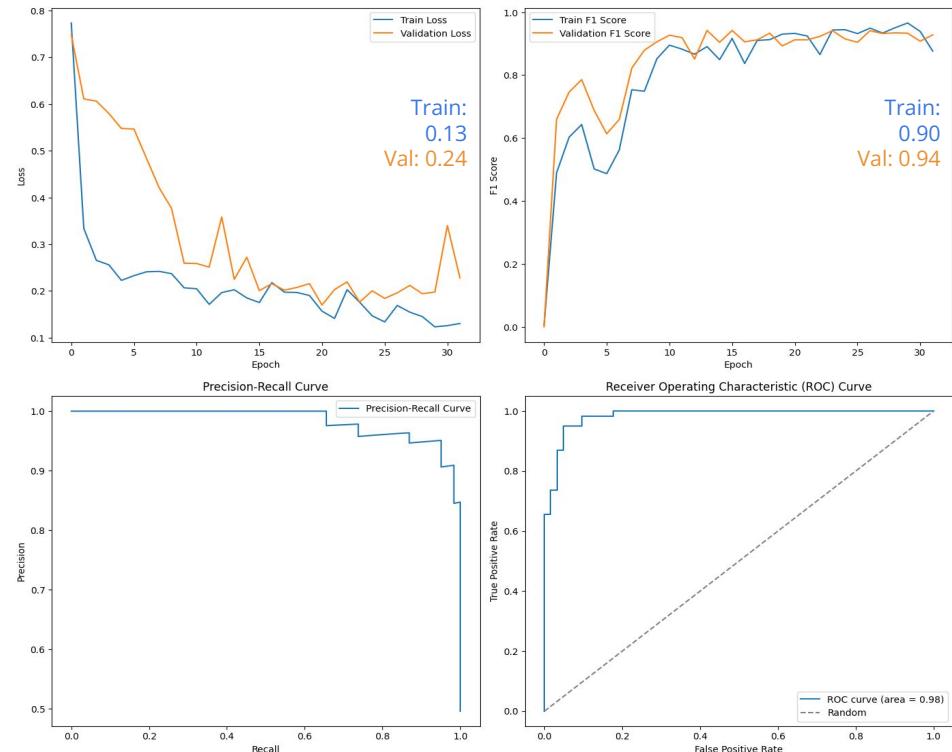
- Model 1:
- 0 - No Advisory
 - 1 - Algae Detected
- Model 2:
- 0 - Caution
 - 1- Danger



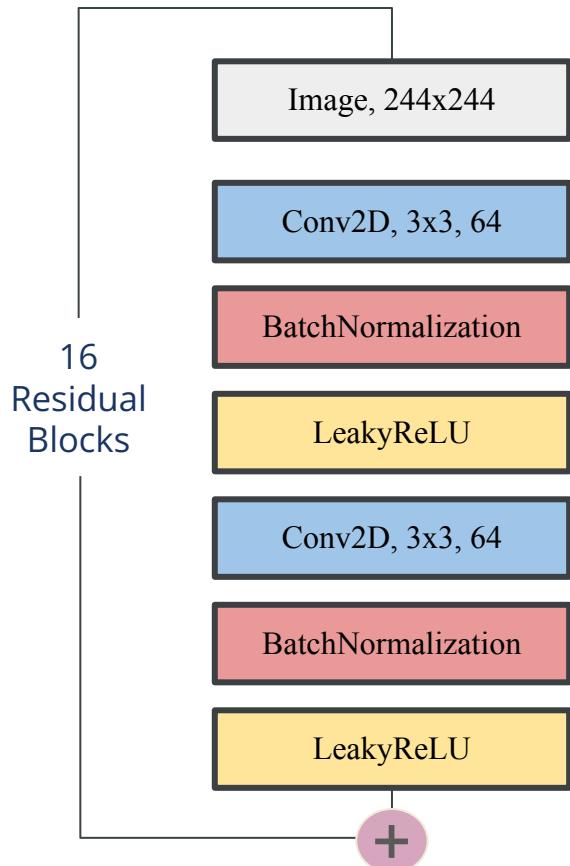
Resnet Model 1 - No Advisory vs Advisory



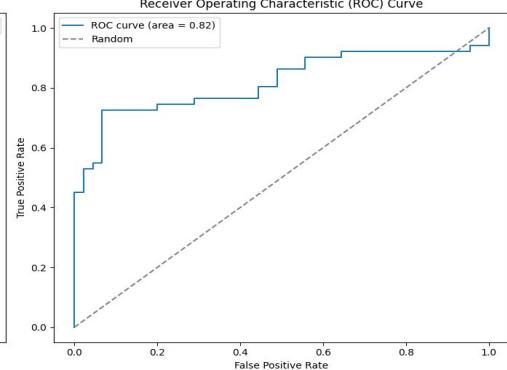
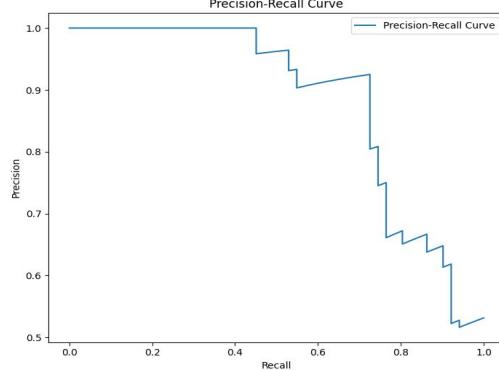
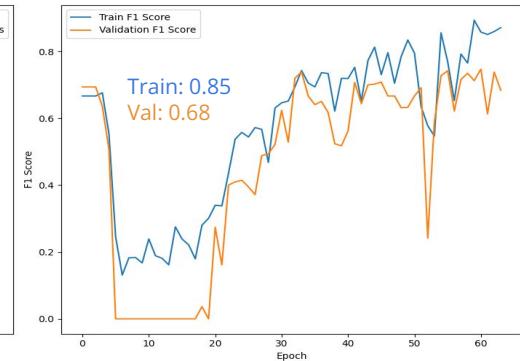
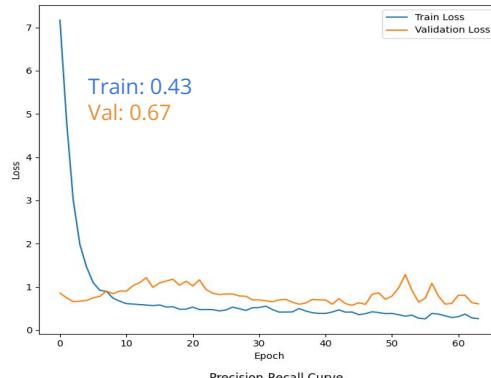
- Healthy loss curve
- F1 score converging
- Performs better than baseline CNN



Resnet Model 2 - Caution vs Danger



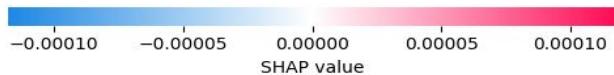
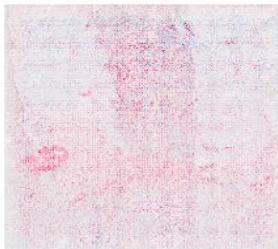
- Plots are not as healthy as model 1 → Suspect due to “caution” and “danger” images being very similar



Model Interpretability

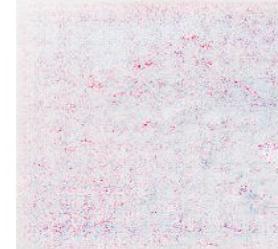
False Negative

True label: 1
Predicted label: 0



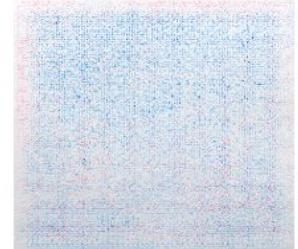
False Positive

True label: 0
Predicted label: 1



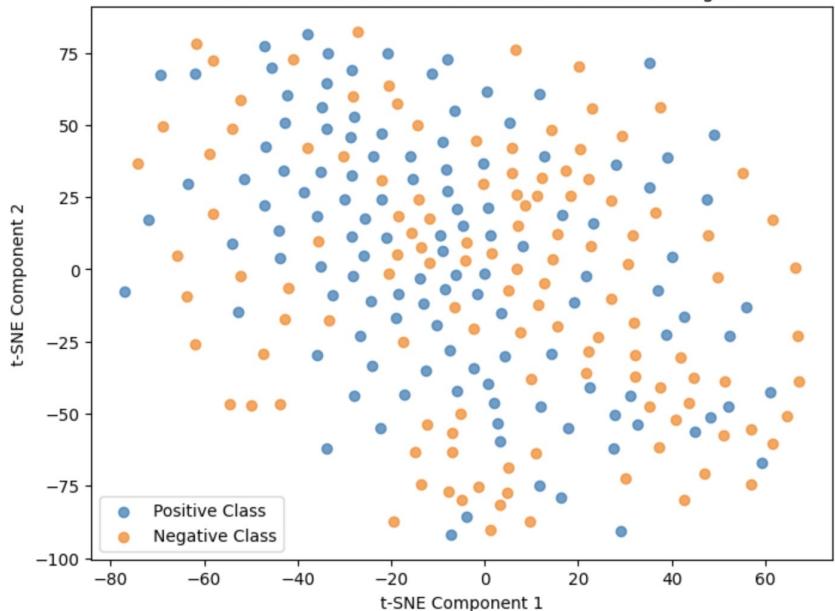
True Negative

True label: 0
Predicted label: 0

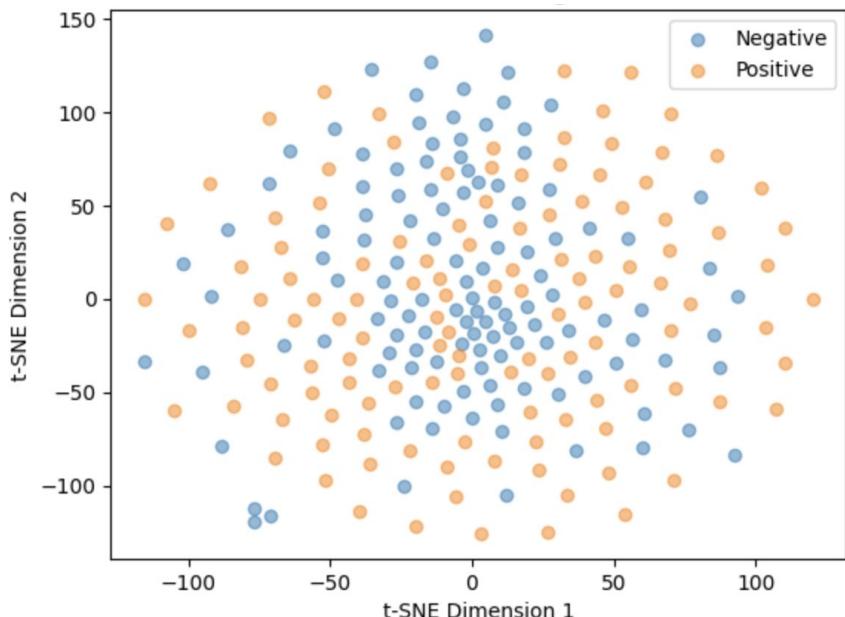


Challenges

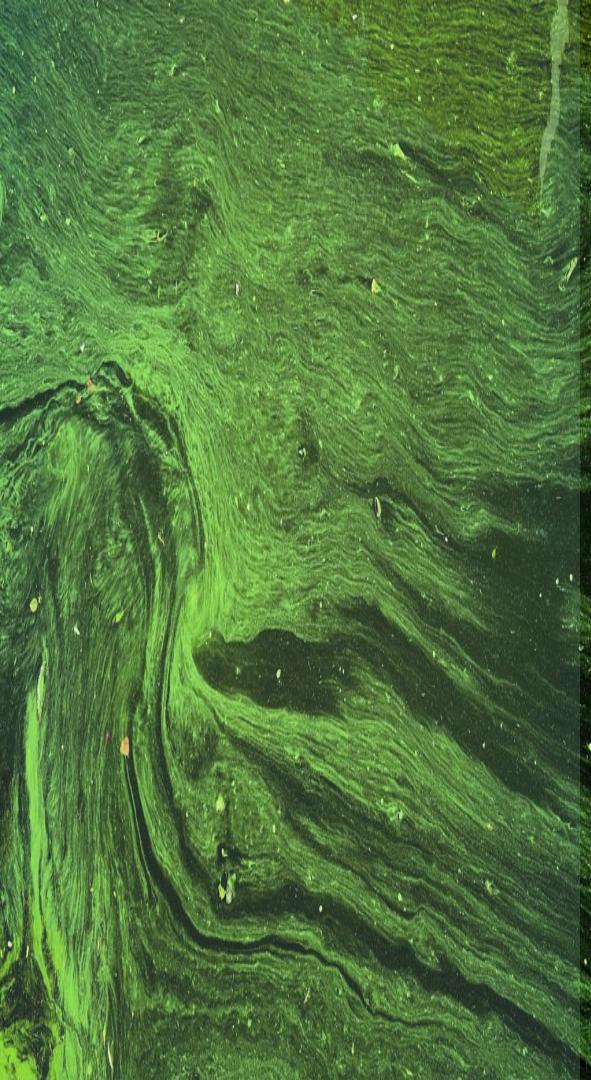
T-SNE Plot of HOG Features



T-SNE Plot of LBP Features



Positive Class = Danger
Negative Class = Caution



Next Steps

- Data Collection
- Further Hyperparameter Tuning
- Further Feature Engineering
- Professional Upload of Microcystin Levels on Streamlit
- Creating Mobile App

Our Mission

1. Detect to defend
2. Reduce the impact of harmful algae exposure
3. Safeguard habitats, preserve water quality, and enhance safety of recreational activities