Modernizing Harmful Algal Bloom Forecasting with Machine Learning & Satellite Imagery

Jesse Cox, Muthumayan Madhayyan, Alexander Caichen, Irina Lee

School of Information, University of California, Berkeley, USA

Abstract

Predicting the location and composition of Harmful Algal Blooms (HABs) allows coastal managers to plan accordingly, protecting drinking water sources, beachgoers, and fisheries. We propose three machine learning models to improve HAB forecasting: first a Deep Fully Connected Neural Network (FCNN) to predict chlorophyll concentration values given a calendar date, a Convolutional Long Short-Term Memory Network (ConvLSTM) for short term chlorophyll concentration predictions, and finally a Convolutional Neural Network (CNN) trained on hyperspectral optical spectrometer data to predict the location of Microcystis, a toxic cyanobacteria. Our solutions focus on the locale of Western Lake Erie due to the presence of workable ground truth data and a well-studied bloom system. Results suggest that modern machine learning methods can produce predictive models for bloom conditions, given a critical mass of quality ground truth data.

1 Introduction

1.1 Problem Statement

Harmful algal blooms are estimated to cause between $10 and $100 million in average annual economic damage [1]. The National Oceanic and Atmospheric Administration (NOAA) produces HAB forecasting maps (Figure 1) to enable coastal managers to mitigate the growing threat of HABs in Western Lake Erie. The maps feature the shape of Lake Erie and a predictive heatmap of "Cyanobacterial Density," an estimation of cyanobacteria cell concentration derived from satellite imagery [2] [3]. The predictions are made by physical models, which employ "a combination of hydrodynamic modeled currents and satellite imagery for bloom location" [4]. Predictions are produced for 1 day and 4 days from the time of posting.

While the United States government has passed several pieces of legislation to combat the negative effects of HABs, [5], and NOAA states that Artificial intelligence is a "key Science & Technology (S&T) Focus Areas that will advance NOAA's mission success" [6], modern data science and machine learning methods are not being directly employed when forecasting HABs in Western Lake Erie.

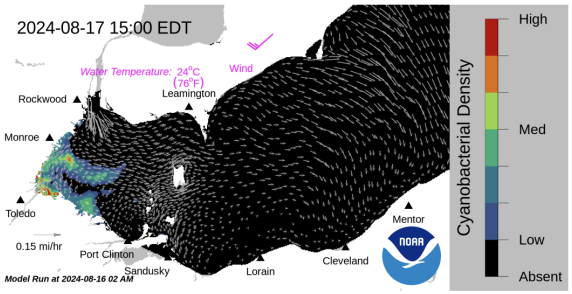


Figure 1: NOAA Archived HAB Forecast Bulletin - 1

1.2 Solutions

We propose three machine learning models to improve HAB forecasting tools for Western Lake Erie. Each of these three models presents a new tool for providing context to HAB forecasting and threat mitigation.

1. A Fully Connected Neural Network (FCNN) that takes a calendar date as input and produces predictive values for chlorophyll concentration for that date. The model is trained on inter- and intra-year trends learned from daily composite satellite imagery from 2017 to present and may be used for long term planning and to make predictions for days when imagery may not be available due to satellite flight path or cloud cover.

2. A Convolutional Long Short Term Memory (ConvLSTM) model trained on 8.5 years of historical Sentinel-3 data, which takes a series of chlorophyll concentration rasters and outputs the next-in-series.

3. An ensemble of Convolutional Neural Networks to predict the presence of toxic cyanobacteria. This model leverages hyperspectral top-of-atmosphere remote sensed reflectance data to detect the spectral signature and spatial patterns of *Microcystis spp*. This model may be used to identify new areas of interest for monitoring *Microcystis* growth, as well as to classify the biological composition of ongoing blooms.

Use cases and tradeoffs between models in spatial, temporal, and spectral resolution are discussed later, followed by challenges in validation as well as suggested guidelines for future improvement.

2 Data

Data from sensors aboard three satellites were used to create the training data for our models, Sentinel-3A, Sentinel-3B, and PACE. All satellite data was accessed with python and "earthaccess," "a community-driven python library that streamlines access to NASA's Earth science data" [7]. After downloading full satellite data files in NetCDF format, top-of-atmosphere remote sensed reflectance values are combined with navigation data to compose a gridded 2D raster at roughly the sensor's native spatial resolution.

Reflectance values from landed areas are not of interest and are masked by default, as are pixels obscured by environmental conditions, like ice or cloud cover. Similarly the satellite's orbit may result in a sensor "miss," where data is unavailable from a large portion of our area of interest. In all cases our models are trained to accommodate for this missing data, so long as an image contains above a threshold percentage of unmasked ("valid") pixels.

Our first two models are exercises in predicting future reflectance or reflectance-derived values, so the satellite data is used as both the training and target variables. For our third model (the CNN), ground truth data was built with water quality monitoring data collected from moored buoys operated by the Great Lakes Environmental Research Laboratory.

2.1 Sentinel-3 OLCI

Sentinel-3A and Sentinel-3B are European Space Agency satellites that "measure Earth's oceans, land, ice and atmosphere to monitor and understand large-scale global dynamics. [They] provide essential information in near-real time for ocean and weather forecasting." Onboard Sentinel-3 satellites, the Ocean and Land Color Instrument (OLCI) is "a visible imaging push-broom radiometer with 21 spectral bands" [8]. The 10 bands with the longest wavelength are not publicly available, leaving 11 spectral bands at 400nm, 412nm, 442nm, 490nm, 510nm, 560nm, 620nm, 665nm, 674nm, 682nm, and 709nm, with roughly 10nm bandwidth. There are also data products derived from the reflectance values, including one representative of chlorophyll-a, which we use in our first two models as indicative of the presence of algae.

OLCI reflectance data is available back to April 26th 2016 and roughly at a frequency of once every 2 days. OLCI spatial resolution is 400m.

2.2 PACE OCI

"PACE is NASA's Plankton, Aerosol, Cloud, ocean Ecosystem mission … PACE is advancing the assessment of ocean health by measuring the distribution of phytoplankton" [9]. PACE's primary sensor is the Ocean Color Instrument (OCI), "a highly advanced optical spectrometer. It enables continuous measurement of light at a finer spectral resolution than previous satellite sensors" [10]. OCI captures 172 reflectance wavelengths, a more than ten times increase over Sentinel-3's OLCI. This comes at a price, however, as OCI has a spatial resolution of roughly 1.2km, meaning that our OCI grids contain 1 pixel for every 9 in our OLCI ones. PACE is a very young satellite and OCI data is only available from April 11 2024 onward at a frequency close to OLCI's.

PACE OCI's hyperspectral imagery presents new opportunities for the advancement of HAB forecasting techniques, backed by research showing that the increased spectral resolution may be used to classify images of different types of algae, especially aided by machine learning methods [11] [12].

2.3 GLERL Water Quality Buoys

The Great Lakes Environmental Research Laboratory (GLERL) and its partners "conduct innovative research on the dynamic environments and ecosystems of the Great Lakes and coastal regions to provide information for resource use and management decisions that lead to safe and sustainable ecosystems, ecosystem services, and human communities" [13].

Since 2012, GLERL has operated a weekly western Lake Erie HAB monitoring program where, during the summer months (bloom season), water is sampled

from buoys and research vessels, before being processed to quantify important HAB indicators, such as water temperature, turbidity, and chlorophyll. Importantly, the samples are measured for the presence of microcystin, a toxin produced by the cyanobacteria (algae) *Microcystis*. Our third model seeks to perform a binary classification to detect the presence of microcystin, and thus *Microcystis*, in order to enhance HAB forecasts, which typically do not include which specific algae is present.

The GLERL microcystin values are available in both "particulate" and "dissolved" form. Simply put, particulate microcystin is that which is contained in cyanobacterial (algal) cells, and dissolved microcystin is free floating and may be present in the water for up to 100 days after the cells containing it have dissolved. Additionally, dissolved microcystin was explained to us to have no optical signature. For these reasons our models were trained to target the presence of particulate microcystin above specific threshold values.

Important particulate microcystin levels are 0.1 µg/L, the minimum detectable concentration given GLERL instruments, 1 µg/L, the WHO maximum recommended concentration for drinking water [14], and 5 and 10 µg/L, both of which serve as warnings against swimming. The distribution of particulate microcystin in our data (Figure 2) is steeply exponential, which presents significant difficulty in predicting conditions with higher concentrations.

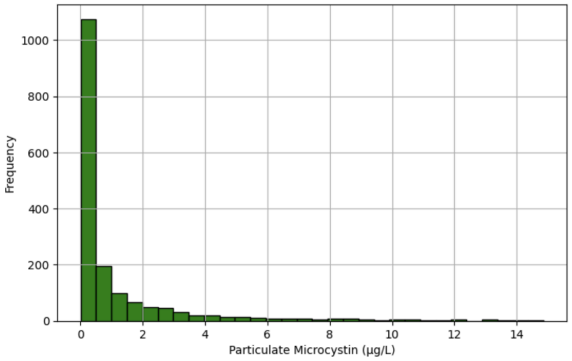


Figure 2: Distribution of Particulate Microcystin

3 Models

3.1 FCNN

Our first model predicts chlorophyll-a concentrations at a day of the year, essentially allowing users to predict algae distributions on days where satellite coverage is unavailable, serving either as fill-in training data for future machine learning models or to guide decision making when data is insufficient to make clear predictions.

3.1.1 Training Data Collection

We use data from Sentinel-3 as the training set for this model, covering Lake Erie conditions from 2017 to present day. For increased accuracy and reduced model size, we only create rasters of chlorophyll-a. Each sample consists of a date corresponding to a 93x163 pixel image of the western region of Lake Erie (Latitude 41.345 to 42.265 and Longitude -83.615 to -81.995). The total dataset contained 1050 samples but was reduced to 620 after removing samples without at least 4000 (26%) non-NaN pixels. Pixel values across the dateset images range from 0.001 to 1000 but generally accrue in the 0 to 10 range.

The input date data was then processed into a numerical format that retains cyclical information by being split into 3 values:

1. Day number in year / Number of days in year
2. $(sin(2 \cdot \pi \cdot dayNumber) + 1) / 2$
3. $-1 \cdot (cos(2 \cdot \pi \cdot dayNumber) + 1) / 2$

Information about the year is not retained since the purpose behind the model is to provide an "average" location based on time of year. We do not want to consider outlier information such as cloud coverage that may be predicted via year, as that does not help predict future years.

Image data, which will serve as the labels for the date encodings, are further processed by applying a mask to indicate areas that should contain water/algae information. The mask is generated using the image

containing the most non-NaN pixels and replaces targeted NaN pixels with 0.001. This mask will reduce the likelihood of the model predicting "0" for lake area pixels and ensure the model consistently predicts an image with a consistent shape. NaN pixel values are also replaced with 0 (Figure 3).

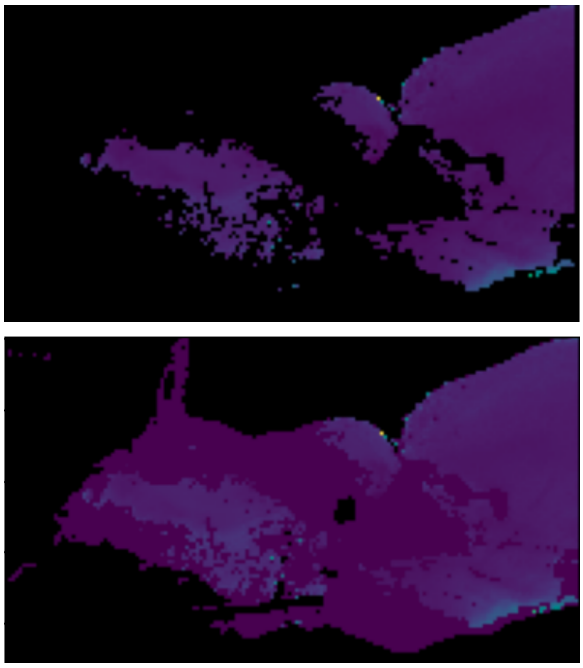


Figure 3: Mask filling from May 22, 2025

Image and date data are then split into training, validation, and testing pairs at a 6:1:1 ratio.

3.1.2 Model Architecture and Training

The model consists of seven dense layers with ReLU activation of increasing size, starting with a layer of size 3 and expanding until a size of 15159, the output of which is then reshaped into a 93x163 image of the Lake Erie region.

Training was done at a rate of 0.001 for 200 epochs in batch sizes of 64, reducing training MSE from 222.686 to 175.841 and validation MSE from 199.033 to 174.122. Final test MSE was 200.749.

Despite the high test MSE, visualizing the model predictions shows the model generalizes well (Figure 5): the expected map displays complex chlorophyll concentration shapes. The Figure's prediction captures some key expected patterns: Higher concentrations near the edges of the lake, especially in Sandusky Bay and the western shore. Other than a protruding bit from the middle of the west coast, other areas in the middle of the lake in the prediction have noticeably lower concentrations.

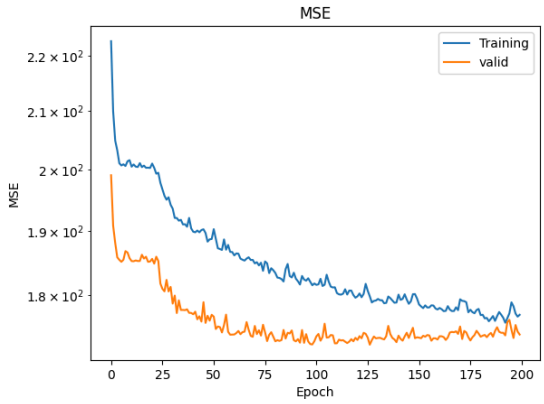


Figure 4: Training loss curve for FCNN

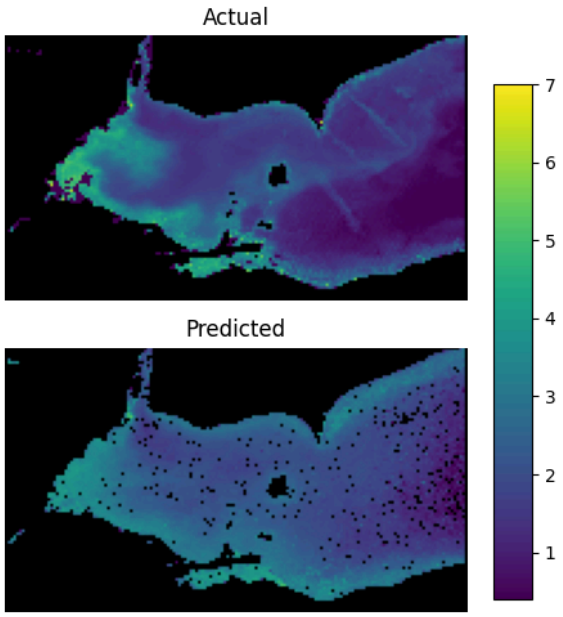


Figure 5: Test set prediction for August 8, 2022

It should be noted that since this model does not take into account certain trends that occur over years, there is a possibility that events such as global warming (that increases temperature and therefore

algae concentrations) over time may quickly cause similar models to be depreciated due to being trained on data that is no longer relevant.

3.2 CNN-LSTM

Our second model is a CNN Long Short-Term Memory (CNN-LSTM) to predict the spatial characteristics of all wavelength reflectances based on a sequence of data from the recent past, enhancing spatial forecasting and capturing temporal dynamics. This architecture integrates convolutional operations with recurrent memory mechanisms, allowing for the modeling of bloom evolution over sequences of satellite imagery. Training and validation are conducted on temporally separated subsets of the dataset to mitigate the risk of temporal leakage and to ensure that model generalization is evaluated across distinct bloom events.

3.2.1 Training Data Collection

We use data from the Sentinel-3 OLCI starting January 2017 through June 2022 as our training set. This captures the temporal variations of reflectance across Lake Erie for about 8.5 years. Data preparation is an important step in achieving a good predictive model. Although the reflectance data has a range of [0,1], almost all of the data is in the range [0, 0.06]. Hence we apply clipping to ensure that outliers do not influence training.

Next we add 3 additional channels to the original 11 channels. Two channels are added as time-markers, which are sine and cosine encodings of the observed day-of-the-year. Another mask channel is added to let the loss functions ignore pixels that represent land, cloud cover, ice etc. The mask channel will not be used in the training dataset.

The typical shape of the training dataset is then:

$$[B, T, H, W, C]$$

where B is the training batch size, T is the number of images in an input sequence, each of which is of size H x W x C. H is the height corresponding to latitude position and W is the width corresponding to the longitude position. C is the number of channels - which is 11 channels + 2 time-channels = 13.

Since the time-channels are in the range [-1, +1], we need to calibrate the reflectance channels to the same [-1, +1] range. Doing so requires us to use a robust scaler.

3.2.2 Model Architecture

The CNN-LSTM model itself contains 2 convolutional layers with 32 filters each, using a kernel size of (3, 3). The output of these convolutional layers is fed as input to a ConvLSTM2D, where the temporal characteristics of the sequence are learned.

We use a sequence size of 5 consecutive frames. Each frame is a snapshot of 11 channels averaged across 3 days. The 6th consecutive frame becomes the target frame. We use a batch size of 4 sequences to iterate through the training process. We also use a custom loss function that computes the MSE between the target and predicted output, ignoring the masked pixels, across all channels. The output of the LSTM layer is then sent to another convolutional layer that then does a spatial reconstruction of the 11 output channels.

Since this is time-series data, we need to ensure that there are no temporal leaks during training. For this purpose we use a validation dataset that starts in Jun 2022 and extends to Oct 2023. We then use a test dataset from Oct 2023 to Aug 2025 to validate how well the prediction engine performs. Our test dataset contains ~116 samples, where each sample is a sequence of 6 frames, where each frame has 11 channels of data. Each channel contains 2 dimensional (H x W) pixels. The 6th frame in the sequence becomes the ground truth for comparison.

The output of the inference contains all the 11 channels. Recall that we need to denormalize the reflectance data. After denormalization, the predicted output can then be compared with the ground truth data.

3.2.3 Results

We could use the most readily available metrics like MAE and MSE. These give an idea of the distance between the true and predicted on the overall frame. However, we found that is too coarse a metric: it does not reflect regional differences, nor does it give a measure of the signal-to-noise ratio. Thus, we take a cue from image comparison metrics such as PSNR and SSIM. These metrics typically compare gray-scaled images. But we extend this to include all 11 channels.

For the entire test dataset, we see PSNR in the range [44, 59] and SSIM in the range [0.874, 0.996] (Figure 6). These results suggest that visual comparison of the image pixels would look fairly similar.

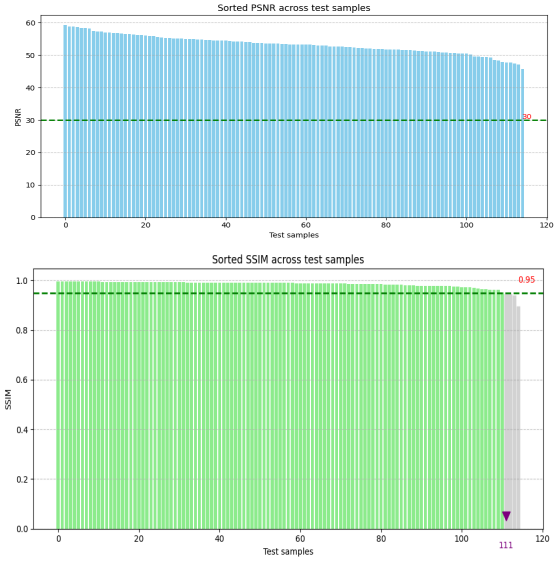


Figure 6: CNN-LSTM Results

3.2.4 Predicting Chlorophyll-a

The same CNN-LSTM model is trained separately on just one channel - the chlorophyll-a readings from Sentinel-3. From a model perspective, everything is identical to the one used for reflectances, except that the shape of the input is not [B, T, H, W, C], where C is just 1 chlorophyll-a channel and 2 day-of-year time channels.

The only other major difference is in the data normalization process. Chla has a wider range [0, 1000]. While most of the data lies with [0, 200], there are some outliers that can skew the spread of data. To guard against this, we use a robust scaler and then a tanh normalization function to get the data between [-1, +1].

We use the same comparison metrics of SSIM and PSNR for this dataset as well. For the entire test dataset, we see PSNR in the range [23.3, 54] and SSIM in the range [0.48, 0.991] (Figure 7).

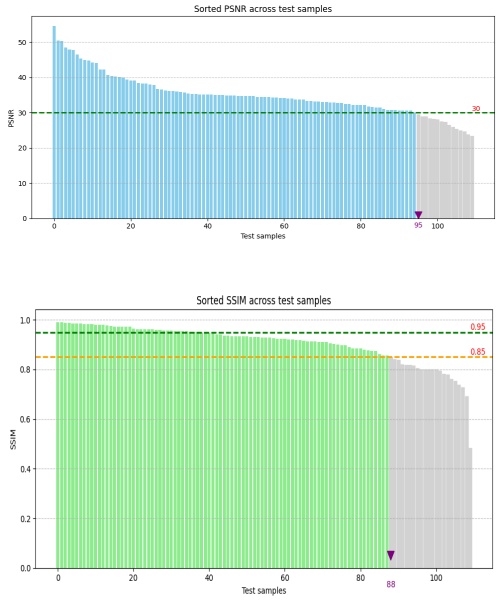


Figure 7: CNN-LSTM Results - Chlorophyll-a

3.3 CNN Ensemble

Our final model is an ensemble of four Convolutional Neural Networks (CNNs) trained to predict the presence of particulate microcystin above a set threshold within a pixel of hyperspectral PACE OCI imagery taken from western Lake Erie. The four CNNs are differentiated by the size in pixels of small image “patches” that form their training data. In order to account for potential bias in the output of any singular model, the final output is the average predicted probability across the four.

3.3.1 Training Data Collection

We begin by using earthaccess to gather all available PACE OCI files that contain data from our area of interest and converting those to two-dimensional rasters of 172 reflectance channels, and an

approximate side length of 1.2km (the native resolution of PACE OCI). We iterate through these rasters where, for each, we search our full list of GLERL water quality observations for data points within a 2-day window (± 1 day) of that image. If a station has one observation present during that window, we extract a "patch" of satellite raster data spatially centered around that observation's latitude and longitude (Figure 8), and use its particulate microcystin value to derive the target variable for its binary classification ('1' if above a selected threshold, '0' if below). The patches have a size of 3x3 pixels, 5x5 pixels, 7x7 pixels, or 9x9 pixels and form the training data for the four discrete CNN models. Patches are only included in the training dataset if the number of pixels that contain valid (non-empty) data across all wavelengths is above a certain threshold percentage. We demonstrate the performance of different thresholds in Figure 10. For our final product, this threshold is set to 50%

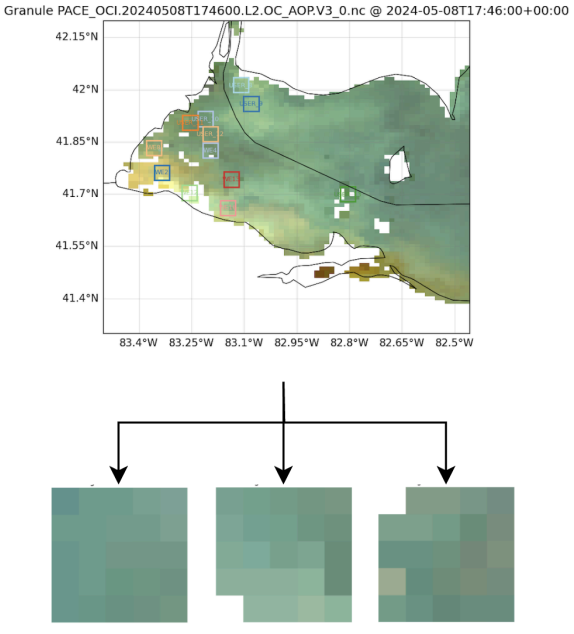


Figure 8: Patch Extraction Illustration

We chose to use square patches with an odd number of pixels on each side to simplify our inference process: our ground truth particulate microcystin observations lie somewhere in the patches' central pixel, while the outlying pixels provide context about the surrounding water. In inference this allows us to "slide" the model over entire unseen rasters, extracting patches and predicting the presence of particulate microcystin in each extracted patch's central pixel.

GLERL monitoring stations are set to record observations only in algal bloom season, roughly from April to October. This fact, combined with the very recent (2024) launch of PACE, drastically reduced our data availability. For example, only ~250 3x3 pixel patches were extracted when the valid-pixel threshold percentage was set to 50%. To accommodate for this we extract patches from random locations across our area of interest from a date range (November to March) in which *Microcystis* is historically not present. This allows us to gather a large number (~1000) of additional patches for which to safely assume the negative binary case (particulate microcystin below the minimum detectable level).

Additionally, after models were initially trained and applied to full rasters to produce maps of the predicted presence of particulate microcystin, we identified a small number (< 20) of pixels from several dates that were consistently and obviously mis-predicted across all models. These were entirely pixels from *very* early in the bloom season (early April) and in the middle of the lake, when and where it is very highly unlikely for Microcystis to be present. Rather than accept these incorrect predictions, we performed a pseudo-labeling process, adding patches around these pixels (labeled negative) to the training data for subsequent models.

Spatial data augmentation techniques were also used to give another large boost to the number of patches contained in the training dataset. Vertical and horizontal flips, and 90, 180, and 270 degree rotations were applied to all patches, while retaining the target variable value for each. This results in a 6X growth in the number of patches to be passed to each model for training. Not surprisingly, this step led to our largest increase in training and validation accuracies, and it seems reasonable to assume that the model began to experience some data leakage in training: as the same patch in different orientations is passed to the model during training, the model may memorize some patch-specific patterns. However, we also saw a significant increase in accuracy and

coherence in the final product (inference), which we believe makes the leakage well worthwhile.

Finally, each patch is accompanied by the "local" per-channel mean and standard deviation of the overall raster from which it was extracted. This provides the model additional context, accommodating differences across rasters that may arise from environmental factors like time of day or relative humidity or satellite operating conditions such as flightpath geometry. All data points are then z-score normalized with "global" per-channel mean and standard deviation taken from all pixels from a large number (100+) of rasters uniformly distributed across the time frame of available data. This ensures all data points are on the same scale, assisting the model in training.

After all these steps, the features to be passed to each model will be of the shape:

$$n \cdot [p^2 + 2] \cdot 172$$

where *n* represents the number of patches that were collected across all sources, *p* represents the height and width of patches in pixels, the *+2* represents the two additional arrays of "local" per-channel raster-specific mean and standard deviation, and finally the *172* reflectance channels.

3.3.2 Model Architecture and Training

All models are trained with the same architecture (Figure 9): two 2D convolutional layers with kernels of size (1, 1) and a number of filters that expands from 32 to 64, followed by a series of 5 Densely connected layers with ReLU activation and a number of neurons contracting from 128 to 1. There are auxiliary layers, such as normalization, reshaping, dropout, and max-pooling throughout the pipeline to aid in feature extraction and generalization. The output layer uses a sigmoid activation to affect binary classification. The models are compiled with an Adam optimizer with learning rate set to 0.0001, binary-crossentropy loss, and AUC of ROC as the primary performance metric.

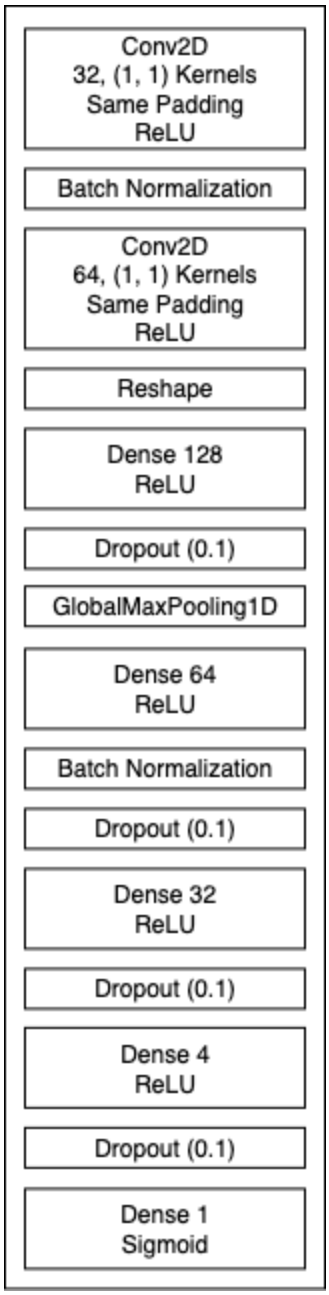


Figure 9: CNN Architecture

Our goal in using layer sizes that expand and then contract is to allow the model to discover more and more new relationships and subrelationships between features for the first half of the pipeline, before narrowing down on which of those relationships are the most important during the second half of the pipeline.

Training and validation accuracy both converge to 100% for every model trained in this regime, likely due to the data leakage discussed above. Test accuracy (F1 like score) similarly converges to 100% in most models, and is always above 95% (Figure 10).

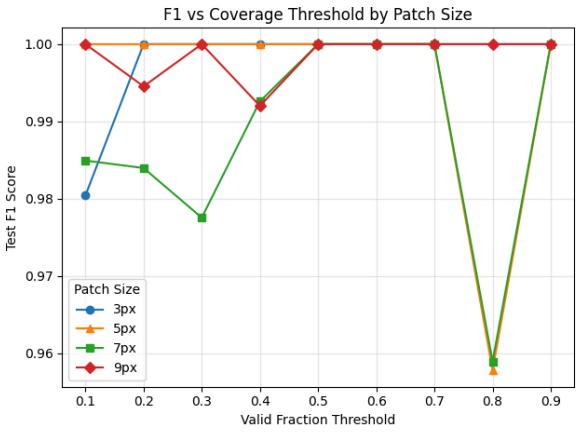


Figure 10: Test Set F1 Accuracies

3.3.3 Output

Each singular model is designed to predict a binary classification of presence of microcystis for the central pixel in a 3x3, 5x5, 7x7, or 9x9 pixel patch. We can then iterate through entire rasters, extracting patches and making binary predictions for each pixel (Figure 11).

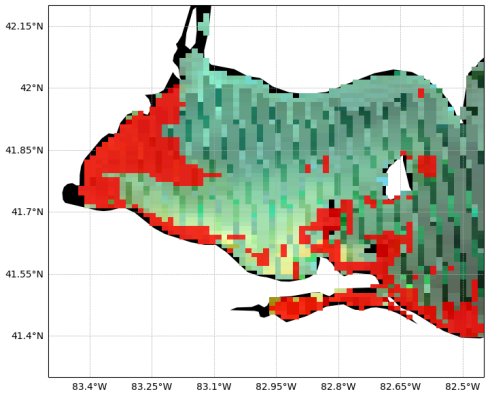
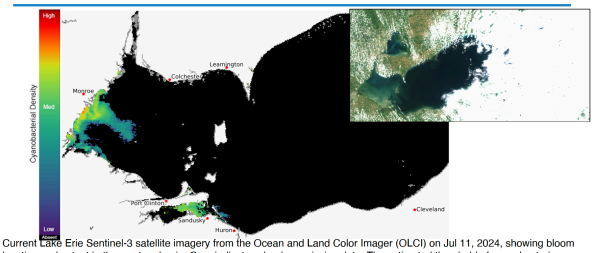


Figure 11: Single Model Output

Our main method for validating the output of any particular model was a visual comparison to NOAA's archived HAB forecasting bulletins (Figure 12), which provide daily plots of cyanobacterial density as a product derived from Sentinel-3 OLCI remote sense reflectances.



Current Lake Erie Sentinel-3 satellite imagery from the Ocean and Land Color Imager (OLCI) on Jul 11, 2024, showing bloom location and extent in the western basin. Grey indicates clouds or missing data. The estimated threshold of cyanobacteria detection is 20,000 cells/mL. Inset shows a truecolor image of the entire lake. Data derived from Copernicus Sentinel-3.

Figure 12: NOAA Archived HAB Forecast Bulletin - 2

It is important to note that these cyanobacterial density values are not necessarily indicative of the presence of *Microcystis*: there are other cyanobacterial algal species in western Lake Erie that may bloom and be detected by the same quantities. We were only able to perform this visual comparison in this situation because we know with confidence that there was a large *Microcystis* bloom in 2024. As such, our output (the prediction of presence of microcystin and thus *Microcystis*) should not be considered a one-to-one analog for cyanobacterial density in all circumstances.

Comparing the results of different models (those trained on different patch sizes) revealed that some models are better than others at predicting certain circumstances. For example, the model trained on 9x9 pixel patches is much more sensitive overall, predicting much larger areas of the positive class, plausibly leading to a higher false positives rates, whereas the model trained on 3x3 pixel patches is more precise, allowing for sharper contours in the shapes it predicts, but potentially resulting in higher false negative rates. Because we do not have access to ground truth for all of Lake Erie, we are unable to determine the overall accuracy of any single prediction map.

To balance these tradeoffs between models, we employed an ensemble method of computing the simple mean of predicted probabilities of particulate microcystin presence in each pixel across all models, and using that mean to set the alpha (transparency) of a layer of red pixels that spans our plot. The result is

a plot where areas of more consistently and confidently predicted microcystin *presence* results in bright red pixels, mixed confidence pixels appear somewhat red, and areas of consistently and confidently predicted microcystin *absence* result in pixels that are completely transparent (or, rather, displays whatever the plot shows behind the prediction mask, whether that be an RGB true color image, plain map, etc.) (Figure 13).

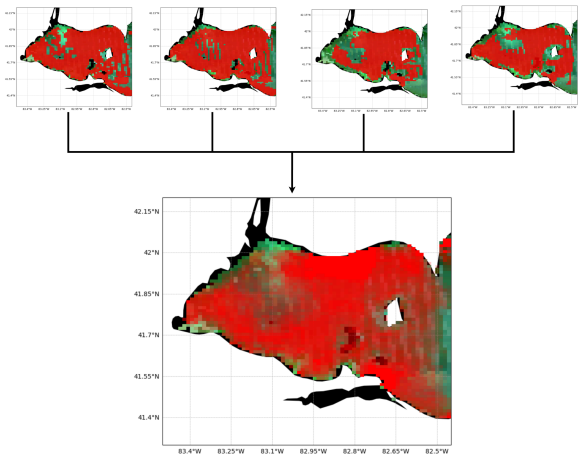


Figure 13: Microcystis Model Ensembling

4 Conclusion

This work demonstrates that modern machine learning methods, ranging from fully connected neural networks to temporal ConvLSTM architectures and hyperspectral CNN ensembles, can enhance HAB forecasting in Western Lake Erie, where there is an adequate amount of ground truth data. By leveraging multi-sensor satellite imagery, ground truth water quality data, and tailored preprocessing pipelines, we have developed models capable of filling observation gaps, capturing short-term bloom dynamics and identifying the likely presence of toxic cyanobacteria *Microcystis*.

Our results show that each modeling approach offers distinct strengths: the FCNN supports long-term planning, or when direct imagery is unavailable, the ConvLSTM captures evolving spatial-temporal patterns in bloom conditions, and the CNN ensemble adds critical biological context by detecting toxin-producing species. Each of these tools can complement existing NOAA physical models, offering improved resolution, adaptability, and specificity.

While the models show strong visual and statistical performance, challenges remain, particularly the need for more extensive hyperspectral training data, improved ground truth coverage across the lake, and safeguards against temporal and spatial data leakage. Future work should focus on integrating these machine learning outputs directly into operational HAB forecasting systems, expanding validation against in situ measurements, and exploring model generalization to other lakes and bloom systems.

By bridging state-of-the-art data science techniques with environmental monitoring, this study points toward a next generation of HAB forecasting that is more accurate, timely, and ecologically informative, aiding coastal managers in protecting water resources, public health, and regional economies.

References

[1] https://coastalscience.noaa.gov/science-areas/habs/assessing-environmental-and-economic-impacts/

[2] https://coastalscience.noaa.gov/science-areas/habs/hab-forecasts/lake-erie/

[3] https://doi.org/10.25923/606t-m243

[4] https://www.sciencedirect.com/science/article/pii/S0034425721004053

[5] https://coastalscience.noaa.gov/science-areas/habs/habhrca/

[6] https://www.noaa.gov/ai/about

[7] https://www.earthdata.nasa.gov/data/tools/earthaccess

[8] https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/olci/

[9] https://pace.oceansciences.org/about.htm

[10] https://pace.oceansciences.org/oci.htm

[11] https://www.youtube.com/watch?v=VKyqYKUNBjY

[12] https://vpresearch.louisiana.edu/news-events/news/20240711/ul-lafayette-researcher-use-hyperspectral-satellites-study-phytoplankton

[13] https://www.glerl.noaa.gov/

[14] https://pmc.ncbi.nlm.nih.gov/articles/PMC6086208/