Monitoring Ocean Health Using AI

A Machine Learning Approach to Satellite-Based Ocean Productivity Assessment

Presented in Applied Machine Learning

2300033083 - Y . Purna Chandra Reddy

2300033191 – K Abhitej Reddy

The Critical Role of Our Oceans

Oceans are fundamental to global homeostasis, acting as the primary regulator of climate and supporting immense biodiversity. Understanding their health is essential for planetary stability.

Climate Regulation

Oceans absorb significant amounts of heat and CO_2.

Ecosystem Foundation

Support diverse marine life, from microscopic algae to large mammals.

The health of phytoplankton is a key metric, as they drive the marine food web and sequester carbon globally.

The Challenge: Dynamic Ocean Productivity

Rapid Variability

Ocean productivity, driven by phytoplankton growth, changes rapidly in response to seasonal shifts, nutrient upwelling, and acute climate variations.

Data Collection Limitations

Traditional ship-based sampling is costly, time-intensive, and provides data limited in spatial and temporal resolution.

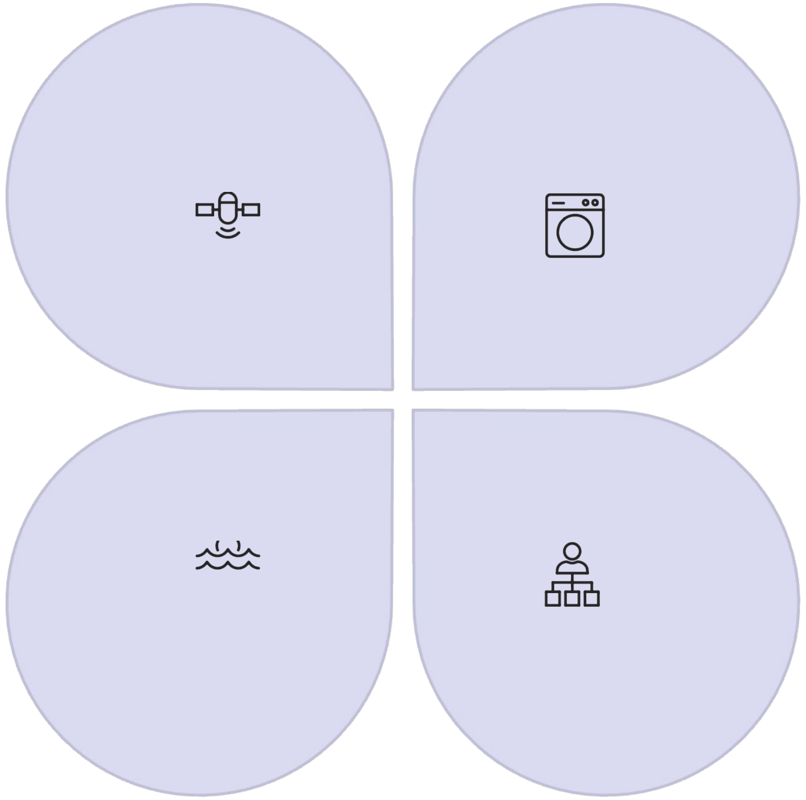
Need for Scalability

Effective global monitoring requires a continuous, scalable, and automated analysis solution utilizing high-resolution satellite imagery.

Project Objective: AI-Driven Ocean Assessment

Analyze Satellite Data

Process vast amounts of remote sensing data to extract key productivity indicators.



Predict Productivity

Use ML models to predict oceanic productivity levels across various spatial scales.

Efficient Monitoring

Establish an efficient, automated system for continuous global ocean health monitoring.

Classify Health Status

Classify regions into actionable categories: Low, Moderate, High, or Harmful.

Feature Engineering from Satellite Remote Sensing

Our predictive model utilizes multi-source, multi-modal satellite data to capture the physical and biological characteristics of the marine environment.

Primary Data Sources

- **NASA MODIS-Aqua:** High-temporal resolution ocean color data.
- **ESA Sentinel-3:** Complementary high-spatial resolution ocean data (OLCI and SLSTR).
- **Platform:** Data aggregation and processing performed via Google Earth Engine (GEE).

Key Predictive Variables (Features)

→ **Chlorophyll-a Concentration (Chl-a):** Direct proxy for phytoplankton biomass.

→ **Sea Surface Temperature (SST):** Influences metabolic rates and stratification.

→ **Salinity:** Indicates ocean currents and freshwater influx.

→ **Photosynthetically Active Radiation (PAR):** Light intensity available for primary production.

Applied Machine Learning Concepts

We utilized robust, interpretable ML techniques suitable for high-dimensional remote sensing data.

Data Preprocessing & Scaling

Handling missing values, temporal alignment, and feature normalization to prepare the data for training.

Random Forest Models

Used for both Regression (predicting continuous Chl-a values) and Classification (categorizing productivity levels). Chosen for non-linearity and robustness.

Rigorous Model Evaluation

Performance assessed using R^2 (coefficient of determination), Mean Absolute Error (MAE), and Classification Accuracy.

End-to-End Methodology: The AI Pipeline

A sequential approach ensured data integrity, model efficacy, and clear interpretability of results.

Collect Raw Data

Acquire multi-band satellite imagery mosaics from GEE.

Clean & Preprocess

Mask clouds/land, harmonize temporal/spatial resolutions, and apply feature scaling.

Feature Extraction

Calculate variables like Chl-a, SST, and derive spatiotemporal features.

Train RF Model

Train Random Forest for regression and classification tasks on historical data subsets.

Predict & Visualize

Apply the trained model to new data and visualize the predicted productivity zones on a map.

Key Results: Prediction Accuracy and Visualization

Model Performance Highlights

93%

Classification Accuracy

Achieved high accuracy in classifying productivity into the four key categories.

0.08

Low MAE Value

Minimal Mean Absolute Error in predicting continuous Chlorophyll-a concentration (mg/m³).

0.88

Strong R^2 Score

Indicates the regression model explains a high percentage of the variance in Chl-a.

Spatial Visualization

Conclusion: AI for a Sustainable Ocean

Scalability Achieved

AI and ML techniques provide a method for large-scale, continuous monitoring of ocean health globally, surpassing the limitations of manual sampling.

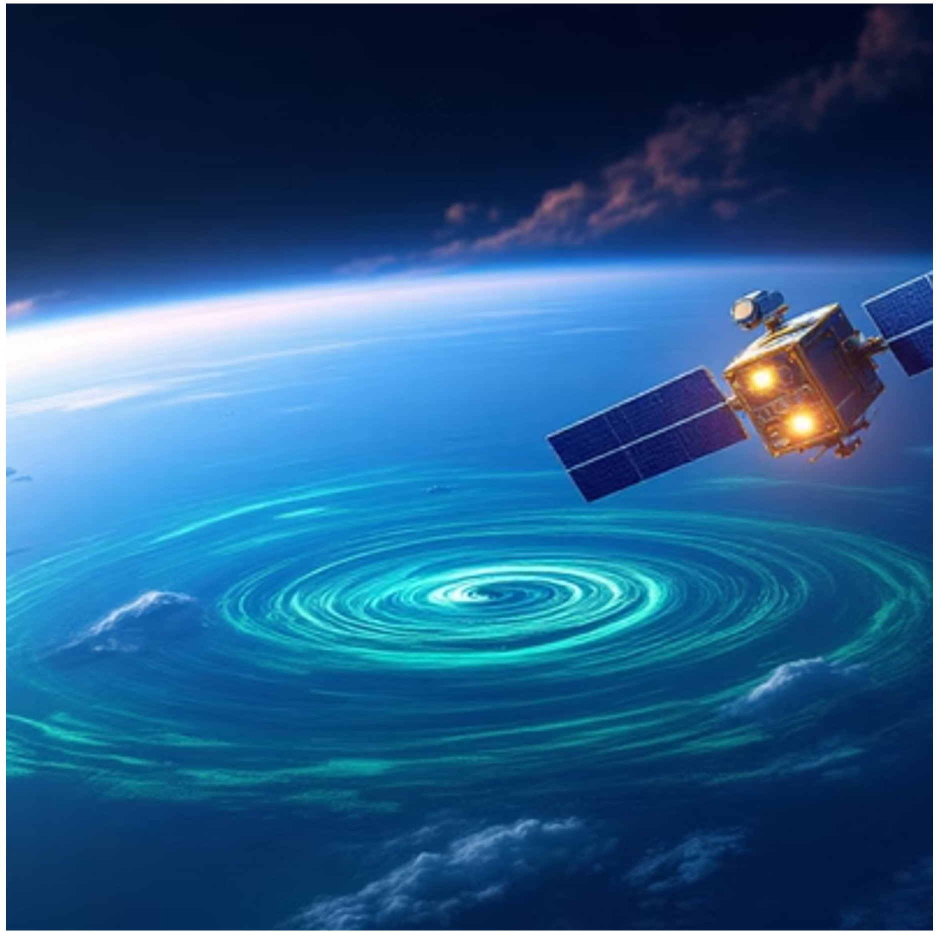
Ecological Insight

Utilizing real satellite data significantly enhances our capacity to perform critical ecological and climate studies, tracking long-term trends.

Practical ML Application

This project successfully demonstrates the utility of applied machine learning in solving complex environmental challenges with actionable results.

Future Work and Outlook



Real-Time Monitoring

Develop the existing model for near real-time operational processing to inform immediate conservation efforts.

Harmful Algal Bloom (HAB) Detection

Focus model expansion on early detection and forecasting of HABs, crucial for fisheries and coastal communities.

Feature Enrichment

Integrate additional data layers, such as ocean current velocities and complex weather patterns, for deeper predictive insights.

Thank You 🌊