Regrading Degrading

Predicting the Current Degradation of Coral Reefs

Nate Leary, Audrey Moessing, Sam Lee, Andrew Goldston, Aidan Quigley

March 2025

1 Introduction

1.1 Research Question

Our project seeks to answer the following research question: *How can machine learning models be used to classify coral reef health status based on image data, and how can these models be leveraged to generate new labeled datasets for monitoring coral degradation?*

Coral reefs are *sub-aqueous calcium carbonate structures* formed by marine invertebrates that play a crucial role in *preserving biodiversity, preventing coastal erosion, and supporting marine economies* through tourism and trade (Spalding et al., 2017; Burke et al., 2011). However, these ecosystems are under severe ecological threat due to rising sea temperatures, ocean acidification, over-exploitation, destructive fishing practices, and marine pollution (Hughes et al., 2017; Pandolfi et al., 2003). Mass coral bleaching events have already been observed in Australia's Great Barrier Reef, serving as an early warning sign of widespread ecological collapse (Hughes et al., 2018). Given these urgent environmental challenges, there is a growing need for scalable and automated solutions to monitor reef degradation. This project aims to address this gap by using machine learning techniques to classify coral health based on image data, with the additional goal of using our trained models to generate new labeled datasets from previously unlabeled sources.

Coral reefs are facing increasing threats from climate change and ocean acidification (Hughes et al., 2017; Pandolfi et al., 2003). The Great Barrier Reef has already experienced mass bleaching events (Hughes et al., 2018).

1.2 Dataset

We will use two primary datasets for our project. The first dataset, sourced from Hugging Face's Coral Health Classification Dataset (Esahit, 2023), consists of labeled coral reef images categorized into three health conditions: *healthy*, *unhealthy*, and *dead*. We will create all of our models using this training data set.

The second dataset consists of unlabeled coral reef images collected by the Yellowfin Surfzone ASV[1] in Majuro, Marshall Islands. These images provide an opportunity to apply our trained models to real-world data, enabling us to generate new labeled datasets for future coral monitoring efforts.

Table 1.2 presents the summary statistics for the dataset, including the mean and standard deviation for each RGB channel, the proportion of white pixels, and the number of observations for each coral health category.

Category	Red	Green	Blue	White Pixel Proportion	Observations
Dead Coral	0.3625 (0.1443)	0.4710 (0.1556)	0.3898 (0.1521)	0.0158 (0.0570)	430
Unhealthy Coral	0.2740 (0.2477)	0.3807 (0.2700)	0.3842 (0.2732)	0.0368 (0.0454)	508
Healthy Coral	0.2554 (0.2521)	0.2935 (0.2328)	0.3206 (0.2632)	0.0091 (0.0272)	661

Table 1: Summary statistics for coral health categories based on rescaled (128 $\times$ 128) image pixel values. We report the mean of each RGB channel, with standard deviations in parentheses. The proportion of white pixels is defined as the percentage of pixels where all RGB channels exceed 200. Red, Green, and Blue values are standardized to the range [0,1] by dividing raw pixel values by 255.

1.3 Machine Learning Methods

To classify coral health and generate new labeled datasets, we will explore three machine learning approaches: Support Vector Machines (SVMs), ordered logistic regression with a ridge penalty, and Convolutional Neural Networks (CNNs).

SVMs are well-suited for structured feature-based classification tasks and have been successfully applied in past studies on coral reef image analysis (Hughes et al., 2017). Ordered logistic regression with ridge penalties provides a useful baseline model for handling ordinal categorical variables while minimizing overfitting. CNNs, on the other hand, are designed for image classification tasks and have demonstrated state-of-the-art performance in recognizing complex visual patterns in ecological datasets (Hughes et al., 2018). We will implement and compare these models, evaluating their accuracy and generalization ability using cross-validation techniques.

Finally, we will apply the best-performing model to the unlabeled Majuro dataset, using its predictions to construct a new labeled dataset for future research.

[1] *Note: We do not currently have access to this data set (approval pending).*

References

Burke, L., Reytar, K., Spalding, M., and Perry, A. (2011). *Reefs at risk revisited.* World Resources Institute.

Esahit (2023). Coral health classification dataset. Accessed: March 15, 2025.

Hughes, T. P., Barnes, M. L., Bellwood, D. R., Cinner, J. E., Cumming, G. S., Jackson, J. B., Kleypas, J., van de Leemput, I. A., Lough, J. M., Morrison, T. H., et al. (2017). Coral reefs in the anthropocene. *Nature*, 546(7656):82–90.

Hughes, T. P., Kerry, J. T., Baird, A. H., Connolly, S. R., Chase, T. J., Dietzel, A., Eakin, M. C., Heron, S. F., Hoey, A. S., Hoogenboom, M. O., et al. (2018). Global warming transforms coral reef assemblages. *Nature*, 556(7702):492–496.

Pandolfi, J. M., Bradbury, R. H., Sala, E., Hughes, T. P., Bjorndal, K. A., Cooke, R. G., McArdle, D., McClenachan, L., Newman, M. J., Paredes, G., et al. (2003). Global trajectories of the long-term decline of coral reef ecosystems. *Science*, 301(5635):955–958.

Spalding, M., Burke, L., Wood, S. A., Ashpole, J., Hutchison, J., and zu Ermgassen, P. (2017). Mapping the global value and distribution of coral reef tourism. *Marine Policy*, 82:104–113.