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.

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.

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

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.

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.

1.4 Machine Learning Pipeline

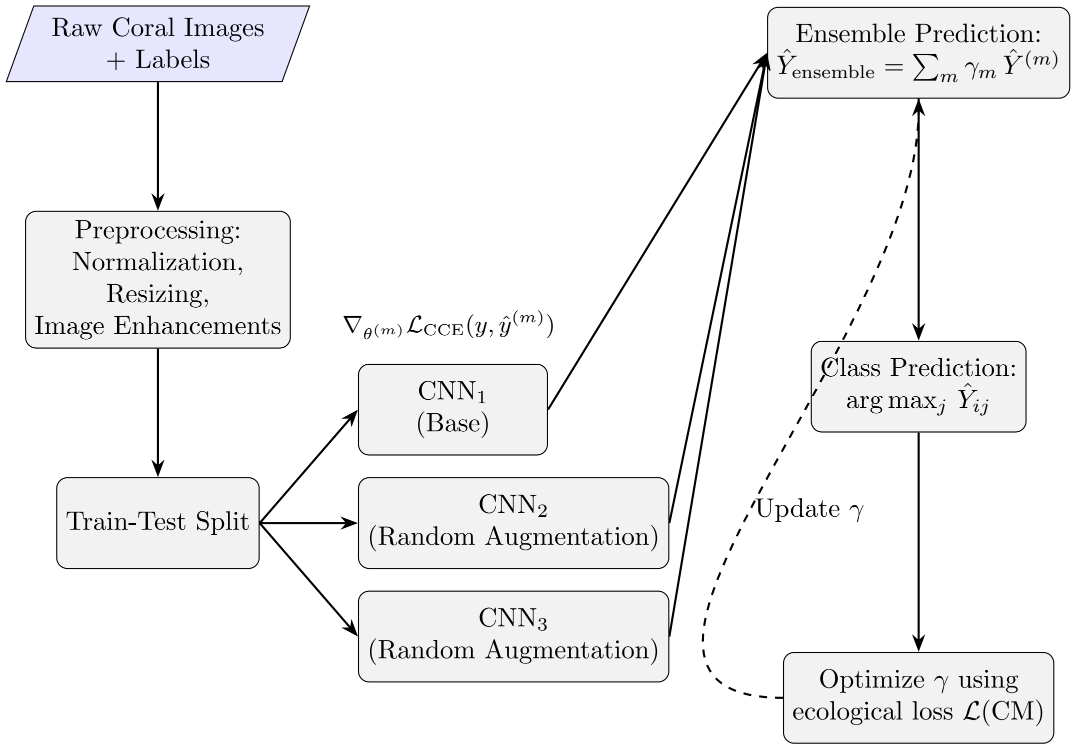

Figure 1: Overview of the Coral Classification Pipeline: raw images are preprocessed, split, and passed through three CNN models with differing augmentation strategies. Their outputs are ensembled using a weighted sum optimized to minimize an ecologically sensitive loss.

1.5 Loss Function

$\Omega =$

True \Predicted	Dead	Unhealthy	Healthy
Dead	ω_{11}	ω_{12}	ω_{13}
Unhealthy	ω_{21}	ω_{22}	ω_{23}
Healthy	ω_{31}	ω_{32}	ω_{33}

$$\mathcal{L}(\text{CM}) = \frac{\sum_{i=1}^{k} \sum_{j=1}^{k} \text{CM}_{ij}\omega_{ij}}{(k^2 - k)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{k} \omega_{ij}} \tag{1}$$

For a given $k \times k$ (3×3 in this case) confusion matrix (CM), We create an *ecologically sensitive* loss function ($L : \Delta_{k^2} \to [0, 1]$) by specifying ω_{ij} s.t. $\omega_{13} > \omega_{31} > \omega_{12} = \omega_{23} > \omega_{21} = \omega_{32} > \omega_{11} = \omega_{22} = \omega_{33} = 0$. The loss function L essentially serves as a *weighted accuracy* function corresponding to what we believe how severe an error we would commit through a false-positive classification (e.g. predicting a healthy coral when it's actually dead). Note that L becomes the normal overall accuracy function for $\omega_{ij} = 1 \, \forall i \neq j$.

For $m = 1, \ldots, M$ convolutional neural networks (CNNs), we train each model using the categorical cross-entropy (CCE) loss. For a single observation with one-hot encoded true label $y = [y_1, \ldots, y_k]$ and predicted class probabilities $\hat{y} = [\hat{y}_1, \ldots, \hat{y}_k]$ where $\hat{y}_j = \frac{e^{z_j}}{\sum_{\ell=1}^{k} e^{z_\ell}}$ and z_j are the pre-softmax logits, the categorical cross-entropy loss[2] is defined as:

[2]The gradient of this loss with respect to the logits z_j is given by:

$$\frac{\partial \mathcal{L}_{\text{CCE}}}{\partial z_j} = \hat{y}_j - y_j$$

$$\mathcal{L}_{\text{CCE}}(y, \hat{y}) = -\sum_{j=1}^{k} y_j \log \hat{y}_j \tag{2}$$

1.6 Ensemble Optimization

Once all $M = 3$ CNN models are trained, we construct an ensemble classifier by forming a convex combination of the individual model predictions. Let $\hat{Y}^{(m)} \in [0,1]^{n \times k}$ denote the predicted probability matrix of model m for n samples and k classes. We define the ensemble prediction matrix as,

$$\hat{Y}_{\text{ensemble}} = \sum_{m=1}^{M} \gamma_m \hat{Y}^{(m)} \tag{3}$$

where $\gamma = (\gamma_1, \ldots, \gamma_M) \in \mathbb{R}^M$ is the vector of ensemble weights subject to the constraint:

$$\sum_{m=1}^{M} \gamma_m = 1$$

We allow γ_m to take on negative values to capture potential anticorrelations between learners, though in practice they typically remain non-negative due to the probabilistic nature of the individual $\hat{Y}^{(m)}$ predictions.

The final predicted class for sample i is then given by:

$$\hat{y}_i = \arg \max_{j \in \{1,\ldots,k\}} \hat{Y}_{\text{ensemble},ij}$$

To determine the optimal weights $\gamma^\star$, we minimize the ecological loss function $L(\text{CM}(\gamma))$, which is defined in terms of the confusion matrix generated from the ensemble predictions:

$$\gamma^\star = \arg \min_{\gamma \in \mathbb{R}^M} L(\text{CM}(\gamma)) \quad \text{subject to} \quad \sum_{m=1}^{M} \gamma_m = 1$$

where $\text{CM}(\gamma)$ is the $k \times k$ confusion matrix comparing the true labels with the ensemble predictions determined by γ. The loss function L is the same ecologically sensitive function previously defined, which weights misclassifications according to their environmental cost.

Optimization via Simulated Annealing

Because the ensemble prediction involves taking an $\arg\max$ over a weighted sum of softmax outputs, the resulting loss surface $L(\text{CM}(\gamma))$ is highly non-convex and non-differentiable with respect to γ. Traditional gradient-based optimization methods such as standard Newton-Rhapson methods are ill-suited to this setting.

To address this, we employ a stochastic global optimization algorithm known as *simulated annealing*[3], implemented via the `dual_annealing` function in `SciPy`. This algorithm combines a probabilistic search over the parameter space with a deterministic local optimizer.

Formally, we solve:

$$\gamma^\star = \arg \min_{\gamma \in \mathbb{R}^M} L(\text{CM}(\gamma)) \quad \text{subject to} \quad \sum_{m=1}^{M} \gamma_m = 1$$

using `dual_annealing`, which is particularly well-suited for this problem due to its robustness to non-smoothness and its ability to handle box constraints on each γ_m. Although we only enforce the simplex

This gradient arises from applying the chain rule to the softmax function and utilizing the fact that $\sum_{j=1}^{k} y_j = 1$. The simplicity of this expression makes CCE not only theoretically well-behaved but also computationally efficient for deep learning optimization.

[3]The algorithm simulates the annealing process in metallurgy, where a material is heated and then slowly cooled to allow it to settle into a minimum energy configuration. At high ẗemperaturel̈evels, the algorithm is permitted to accept uphill moves with the objective of escaping local minima. As the temperature decreases, the algorithm becomes increasingly greedy, converging toward a local or global minimum.

constraint $\sum_m \gamma_m = 1$ (and allow γ_m to take on negative values), we normalize γ during optimization to remain on the simplex manifold.

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.