Eindhoven University
of Technology (TU/e)
Mathematics & Computer Science

Quantifying Coral Health with Segmentation Derived Features

Capstone Data Challenge - JBG060

Authors: David Mandado, Marc Boglioni, Pelayo Davara, Timo Wijnen, David Ilisei
Supervision: Michael Cevaal
Date: Quartile 1 2025
Topic Researching coral health through image feature extraction

Abstract

Coral bleaching threatens reef ecosystems, but many automated assessments rely on opaque, compute-heavy classifiers. We use ML only for coral/non-coral segmentation and infer condition from interpretable texture and color/whiteness features computed *within* masks. A U-Net with an EfficientNet-B0 encoder provides masks (mIoU ≈ 0.67). On top, we demonstrate (A) a linear, regression-based health index and (B) percentile-matched exemplar retrieval for qualitative validation. Whiteness and texture behave as expected with the bleaching proxy, supporting their use as transparent indicators. The index+exemplars design offers explainable monitoring on stakeholder data, the main limitations are data volume and lack of ordinal severity labels.

Contents

1 Context, Problem Background, and Motivation

Coral reefs cover only $\sim$1% of the ocean yet support $\geq$25% of marine life (NOAA, n.d.). Rising temperatures and acidification increase bleaching prevalence (Van Woesik et al., 2022). Manual grading is slow and often binary (bleached vs. not). Instead, coral is segmented and texture and color/whiteness are quantified *within-mask* to obtain a graded signal without black-box severity models (Figure 1).

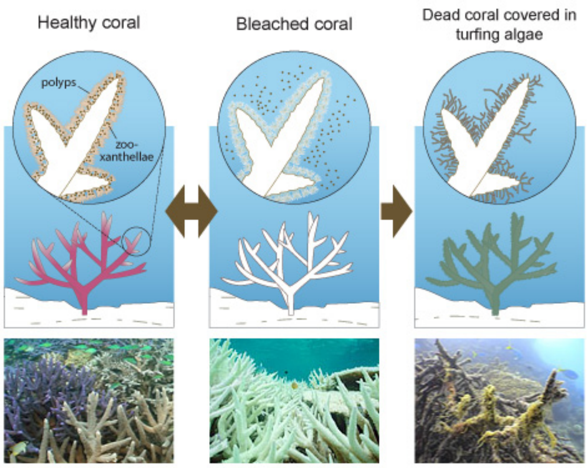

Figure 1: Healthy/Bleached/Dead coral (Valinsky, 2015)

2 Business and Data Understanding

2.1 Prior Work

Most image-based reef studies deliver either patch-level *classification* (e.g., CoralNet) or semantic *segmentation* for percent cover, yielding discrete outputs rather than within-coral condition (Beijbom et al., 2012; CoralNet, n.d.). In contrast, comparatively few works extract literature-grounded, interpretable features (e.g., texture, color characteristics) directly from images to quantify health states. Field practice already treats health as graded: the CoralWatch colour score provides an ordinal pigmentation scale for bleaching severity (CoralWatch, The University of Queensland, n.d.), and ecological surveys often report continuous/ordinal severity (e.g., percent colony bleached) (Van Woesik et al., 2022). Our approach explicitly targets this gap by mapping within-mask features to a graded health indication, beyond binary bleached or not-bleached.

2.2 Data Sources and Properties

This study uses Reef Support's **Bleaching** and **Benthic** datasets. The former provides bleached vs. non-bleached masks, the latter contains hard vs. soft coral masks. Combined, they yield coral vs. non-coral masks for 3,500 images across multiple reef sites. The imagery spans diverse conditions (lighting, turbidity, and color casts) as seen in Figure 2 introducing domain shift that is addressed through preprocessing and site-stratified splits.

Additional public datasets were evaluated but excluded due to incompatible or unreliable labels. Consequently, our labels indicate presence/absence of bleaching rather than graded severity, motivating our feature-based health estimation instead of supervised severity prediction.

These data directly serve the stakeholder's needs: they are operationally available to Reef Support, representative of their deployment settings, and sufficient to train robust segmentation and to compute interpretable within-mask features.

2.3 Risks and Mitigations

Cross-site domain shift: Varying turbidity, illumination, and color casts can bias both masks and feature magnitudes. *Plan:* normalization and site-aware evaluation.

Label scope (binary, not graded): Available labels don't encode severity. *Plan:* use interpretable features (whiteness/texture)

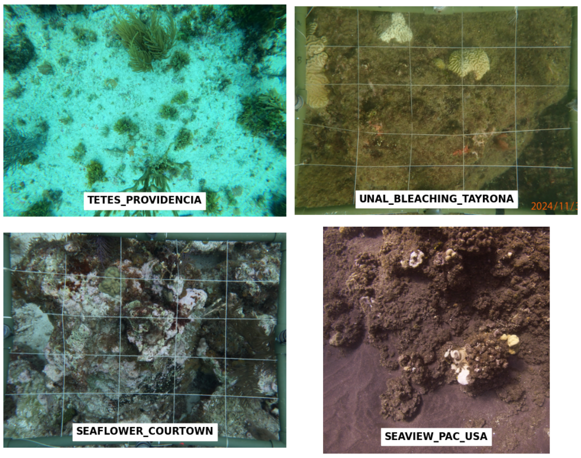

Figure 2: Comparison of sites in the dataset

to approximate graded health, validate directionality.

Mask quality and imbalance: Noisy masks and low coral coverage increase variance. *Plan:* basic mask sanity checks, report coverage effects.

Spurious whiteness: Sand/light may mimic bleaching. *Plan:* restrict to coral masks, include texture/morphology cues.

3 Research Objective

3.1 Main Goal

Achieve *graded* assessment of coral condition from photo-quadrats using a transparent, low-compute pipeline that (i) relies on ML only for coral/non-coral masks and (ii) infers health from interpretable, literature-motivated features. This goal follows directly from the context (need for scalable, explainable monitoring) and the available data (binary bleaching and benthic masks, Section 2).

3.2 Objectives

- **Reliable segmentation:** Train and evaluate a model that isolates coral from background (metric: mIoU).
- **Interpretable features:** Find features that applied to masks, can be identified to correlate with bleaching. As well as having sufficient features to achieve a score for texture and color/whiteness per image.
- **Two complementary health outputs:**
  (a) *Regression-based health index* aggregating standardized features into a graded score.
  (b) *Percentile-matched exemplars* retrieving images with similar joint feature percentiles to support qualitative validation and explainability.

3.3 Assumptions

- *Feature–health linkage:* After per-site normalization, increased whiteness (reduced pigmentation) and reduced texture complexity indicate higher bleaching severity, assumption based on literature research.
- *Mask fidelity:* Segmentation masks are sufficiently accurate for within-mask statistics (e.g., exclude images with coral coverage $<$ X% or with mask QC flags).
- *Operational constraints:* Due to limited data available and computational-resources focus is on transparency and rigorous feature findings.

3.4 Measurable Adequacy Criteria

- **Segmentation quality:** Mean IoU $\geq$ *0.65*.
- **Index validity:** Appropriate reasoning that computed index and the assumption used are justified by feature correlation and/or literature.
- **Stakeholder:** Feature extraction results and approaches (a) and (b) must demonstrate whether feature analysis from images is successful and viable for further research or proves inadequate to approach coral health.

4 Approach

4.1 Pipeline Overview

Our pipeline (see Figure 3) begins with coral background *segmentation*. U-Net variants where benchmarked and encoder/backbone choices while tuning preprocessing, augmentation, loss functions (e.g., BCE+Dice), and key hyperparameters to maximize mIoU on site-held-out data. Using the resulting masks, compute *within-mask* features (color and texture designed to reflect tissue condition). After, evaluate two complementary routes: (a) a transparent, regression-based health index aggregating standardized features, and (b) percentile-matched exemplar retrieval that surfaces images with similar joint feature percentiles for qualitative validation. Together, these steps provide graded, interpretable health indications beyond binary bleaching.

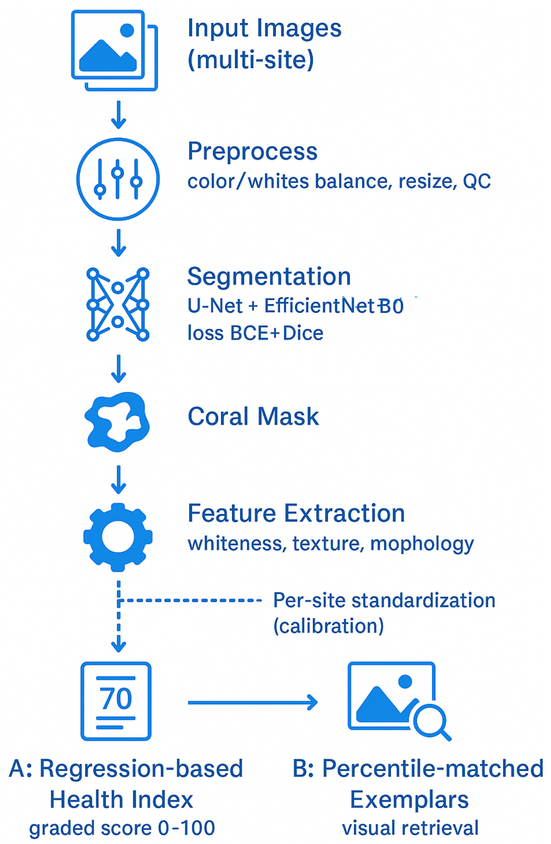

Figure 3: Pipeline Process Diagram

4.2 Preprocessing

To keep the focus on *interpretable health estimation*, a lean preprocessing stack was used and only essential training-time augmentations. Concretely:

(1) Mask QC and filtering, removed empty/invalid masks and discarded images with trivially small coral regions to avoid unstable within-mask statistics (see 4).

(2) Uniform resizing, all images and masks were resized to 512 × 512, which preserved colony morphology while reducing compute for both training and feature extraction (see 5).

(3) Site-stratified splits-train/validation/test partitions were stratified by site to prevent leakage and to measure generalization under domain shift (seen in Figure 2 before).

4.3 Segmentation

Model coral-background *segmentation* with a U-Net decoder and an EfficientNet-B0 encoder. Inputs are $3 \times 512 \times 512$ RGB images, outputs are $1 \times 512 \times 512$ probability masks. Training uses Reef Support's Benthic and Bleaching datasets after mask QC (Section 2) and resizing. The loss is a weighted sum of BCE and Dice (0.7/0.3). Optimize with AdamW (lr = *3e-4*, weight decay = *1e-5*) and ReduceLROnPlateau on validation IoU. Augmentations are conservative (horizontal/vertical flips, small rotations, brightness/contrast and saturation jitter).

Evaluation follows a *site-stratified* protocol reporting mean IoU. The best checkpoint attains **mIoU** $\approx$ **0.67** on held-out sites, with expected drops in turbid/low-contrast scenes (see Appendix A for Optuna hyperparameter optimization results).

4.4 Feature Extraction

All features are computed *within the coral mask* to avoid background contamination.

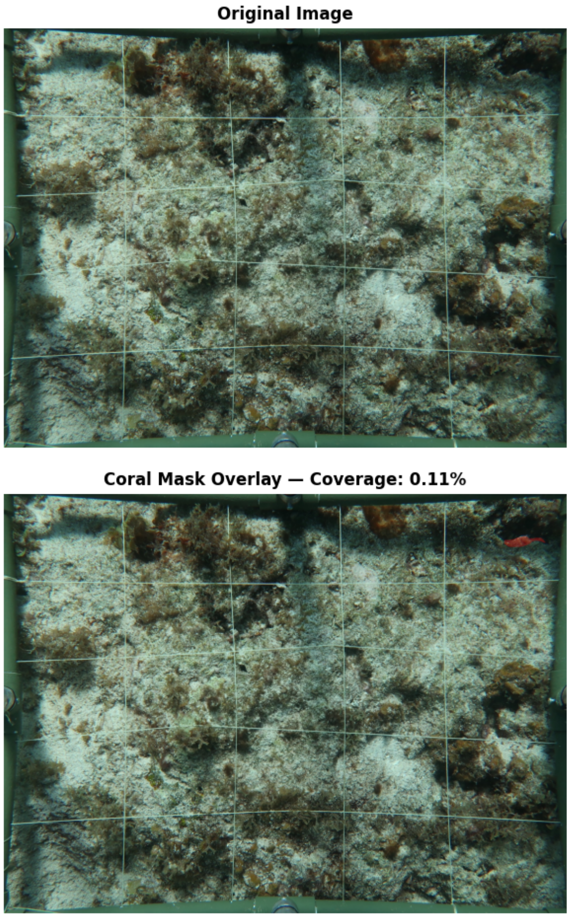

Figure 4: Small mask example removed

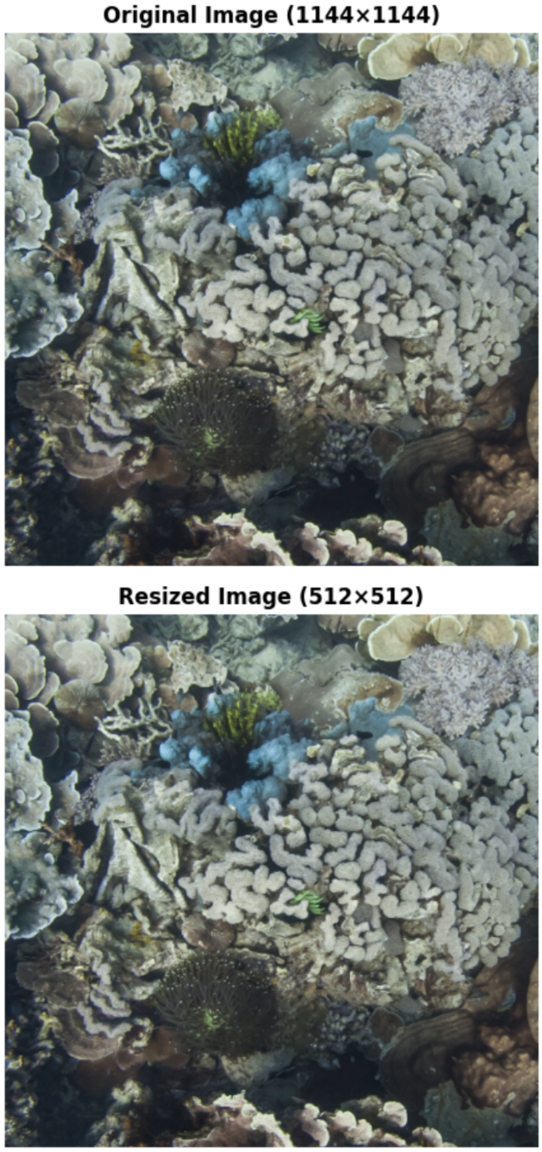

Figure 5: Resized image comparison

Below are the final features used, with brief rationale.

4.4.1 Texture / Structure (3)

- **Laplacian variance (grayscale).** Quantifies high-frequency detail, smoother (bleached) tissue generally shows lower values.
- **LBP variance (grayscale, P=8, R=1, uniform).** Captures micro-texture richness, healthier, rugose tissue tends to yield higher variance.
- **GLCM correlation (dist=1; angles $0°, 45°, 90°, 135°$; averaged).** Measures grey-level dependence, changes reflect altered micro-organization under bleaching.

As justification to the texture features that used grayscale (Laplacian and LBP), both features showed the same behavior. Highly biased distribution of the feature with plain colors (see Figure 6), which was balanced by applying grayscale to the coral areas on which the features were computed (see Figure 7).

4.4.2 Color / Whiteness (6)

- **Raw red intensity.** Proxy for pigmentation, bleaching often reduces red-channel contribution.
- **Albedo (illumination compensated**

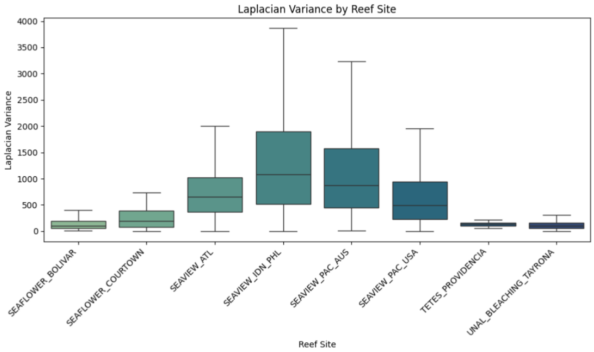

Figure 6: Laplacian variance across reef sites

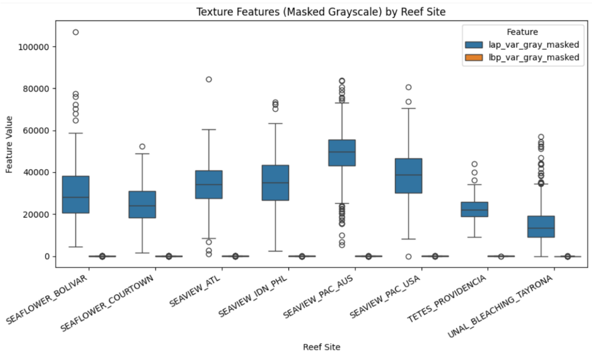

Figure 7: Laplacian variance across reefs after grayscale

brightness). Approximates intrinsic surface reflectance, increases as tissue whitens.

- **Luminance (perceptual brightness).** Overall lightness, bleaching raises apparent luminance as pigments are lost.
- **Saturation (HSV).** Pigment loss lowers saturation, useful to separate pale tissue from normally colored colonies.
- **Median raw red (robust central tendency).** Reduces sensitivity to outliers/speculars while preserving pigment signal.
- **Median luminance (robust brightness).** Stabilizes brightness estimation under highlights/backscatter typical in underwater scenes.

scriptors (e.g., overlapping GLCM metrics) are flagged and pruned before modeling.

4.5 Approach A: Regression Health Score

Two interpretable sub-scores from site-standardized features: a *texture score* $S_{\text{tex}} = 0.121\,\text{Lap} + 0.067\,\text{LBP} + 0.469\,\text{GLCM}$ and a *whiteness score* $S_{\text{white}} = -0.124\,\text{Red} + 0.154\,\text{Alb} + 0.198\,\text{Lum} + 0.289\,\text{Sat}$ (see Table 1). The final index is a convex combination $H = w_{\text{tex}} S_{\text{tex}} + w_{\text{white}} S_{\text{white}}$ (Table 2), scaled to $[0, 100]$ using min–max on the validation split. Signs match domain intuition (paleness $\uparrow \Rightarrow$ health $\downarrow$; texture $\uparrow \Rightarrow$ health $\uparrow$), keeping the model transparent and easy to audit.

4.6 Approach B: Percentile-Matched Exemplars

For this approach a code block finds, in the bleaching dataset from Reef Support, the images in which all 9 selected features are as close as possible to the same percentile. all of the features and standardized for both approaches. Thus, when computing and image at the 40th percentile, that image is the best match in the dataset where all features are as close to the 40th percentile in its own feature's distribution. This approach is a simple demo to what could be identified as middle points in health levels, but do not account for the difference in weights that each features can have in real life coral.

4.7 Generalization and Assumption Handling

- **Mask quality:** Empty/tiny/invalid masks removed, coverage tracked to flag unstable estimates.
- **Interpretation:** Assumed that higher paleness and lower micro-texture reflect bleaching, exemplars are used as a human cross-check. These are validated by literature but still reduce to an assumption.

5 Evaluation Methodology

Segmentation: mIoU. **Features:** distributions by site; prune highly collinear descriptors. **Index:** train/val split, and rank correlation vs. % bleached, directionality sanity checks. **Exemplars:** stakeholder face-validity (usefulness/explanation), and average percentile distance. All measures map to adequacy criteria (Section 3.4).

6 Results

Note: All figures in this section have extended interpretations in Appendix B.

6.1 Segmentation

Under the site-stratified protocol, the final U-Net (EffNet-B0) achieves **mIoU ≈ 0.67** on held-out sites, with performance varying by turbidity/lighting. Qualitative inspection confirms good boundary adherence on shallow, well-lit images and under segmentation in backscatter or very low contrast.

6.2 Feature Relationships

Whiteness block. The pairplot in Figure 11 shows a coherent paleness signal: albedo and luminance vary together, while saturation and raw red move oppositely. Points against the last column indicate that higher paleness aligns with higher bleaching.

Texture vs. bleaching. Figure 12 illustrates the expected pattern: as colonies bleach, high-frequency detail and micro-texture diminish (lower Laplacian/LBP; higher regularity in GLCM correlation).

Whiteness vs. bleaching. Figure 13 shows a complementary trend on pigmentation: brighter/whiter and less saturated tissue corresponds to higher bleaching.

Takeaway. Color/whiteness captures *pigment loss*, texture captures *surface smoothing/regularity*. The two families are complementary rather than interchangeable, which motivates using both in the health index and exemplar demos.

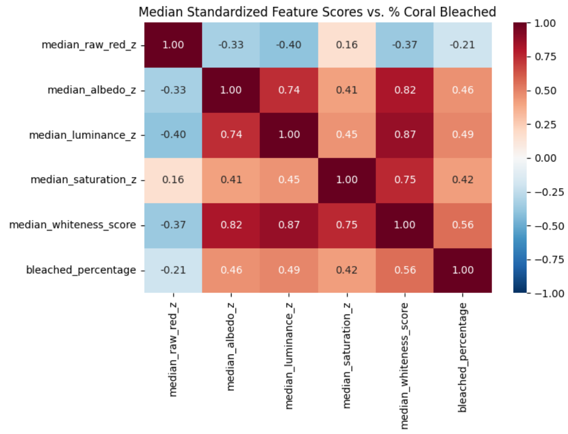

Figure 8: Correlation matrix for color/whiteness features and % bleached.

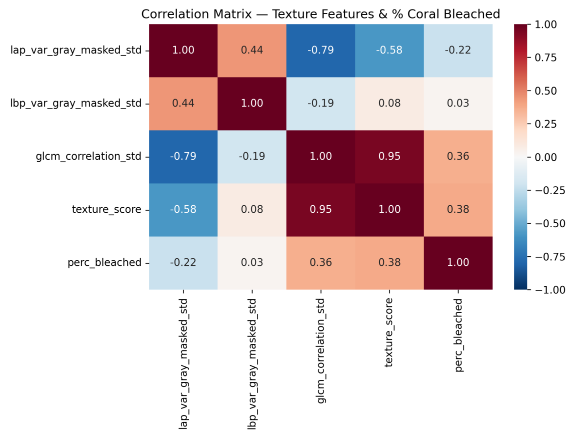

Figure 9: Correlation matrix for texture features and % bleached.

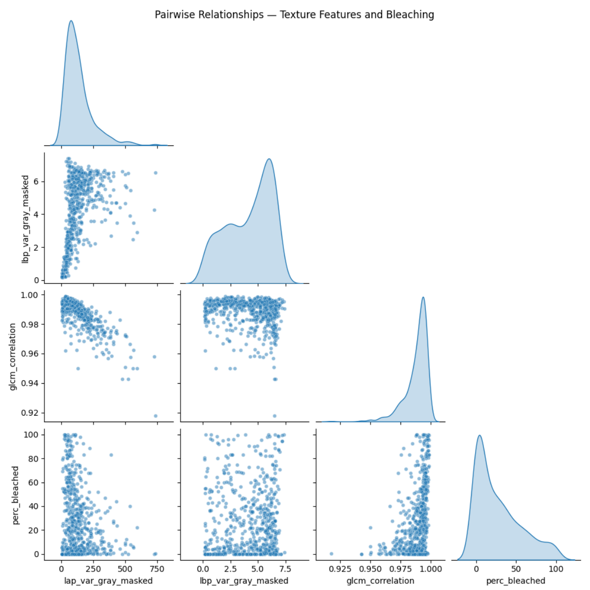

Figure 10: Scatterplot matrix: Laplacian variance, LBP variance, GLCM correlation vs. percent bleached.

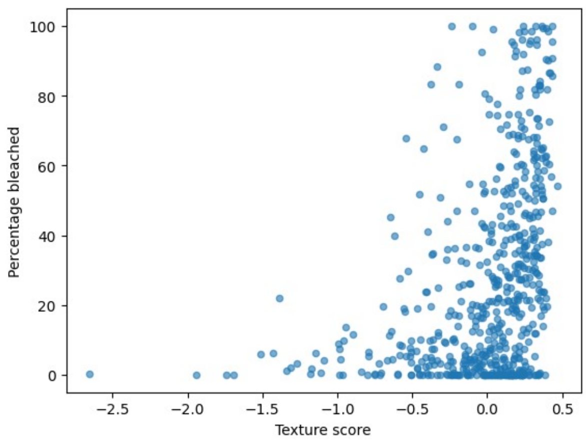

Figure 12: Texture features vs. percent bleached (representative scatter).

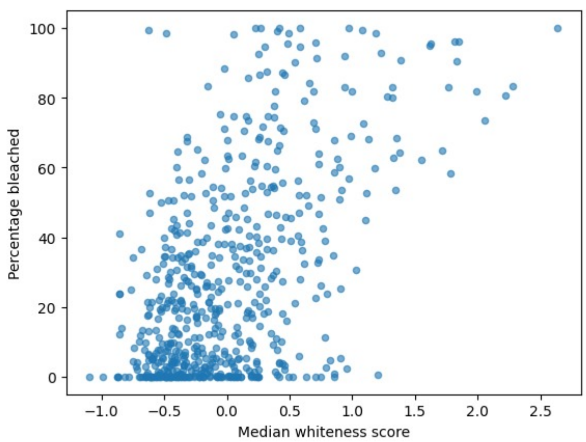

Figure 13: Whiteness feature vs. percent bleached (representative scatter).

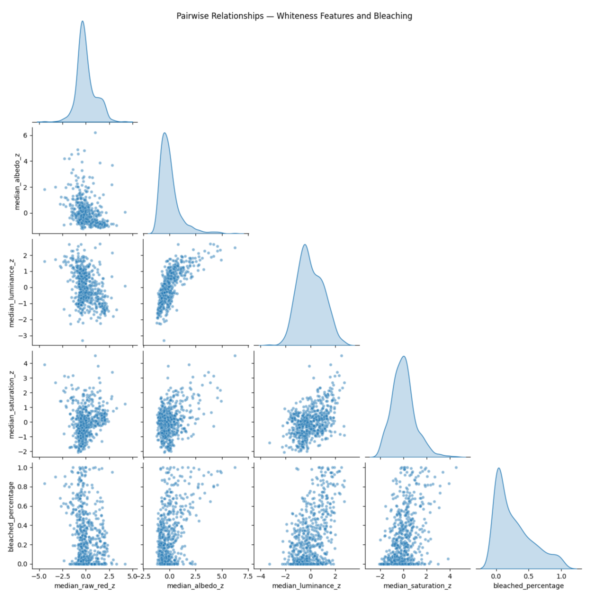

Figure 11: Pairwise relationships among whiteness features (red, albedo, luminance, saturation) and percent bleached.

6.3 Approach A Demo

For this approach, a demo of the method is shown. Table 1 shown the weights for each feature for the texture and whiteness score of a coral region, and Table 2 the weights of texture and whiteness overall for a final score. Figure 14 shows how our features differentiate health in corals automatically

Table 1: Final regression weights for the two interpretable scores (higher = healthier).

Score	Feature	Weight
Texture	Laplacian variance	0.12
Texture	LBP variance	0.07
Texture	GLCM correlation	0.47
Whiteness	Raw red	−0.12
Whiteness	Albedo	0.15
Whiteness	Luminance	0.20
Whiteness	Saturation	0.29

Table 2: Final score composition

Component	Weight
Texture score	0.25
Whiteness score	0.50

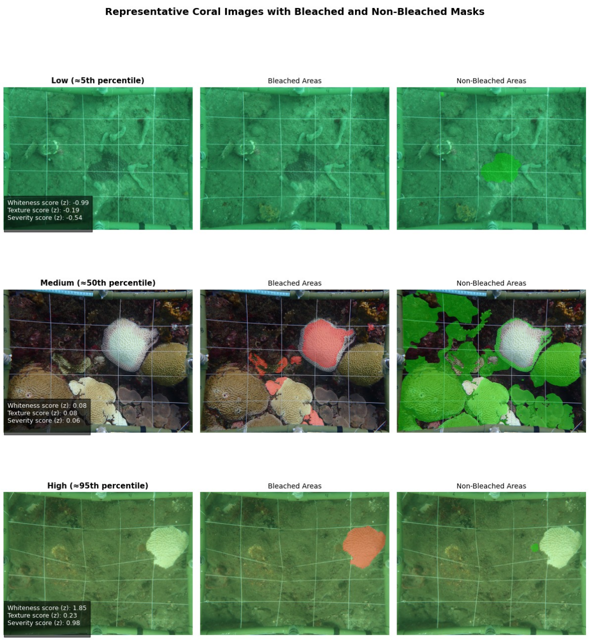

Figure 14: Approach A example

6.4 Approach B Demo

As the method for approach B was explained before, an example of this method will be used as result, composed of 5th, 40th, 60th and 95th percentiles to capture middle points of health (see 15 and 16).

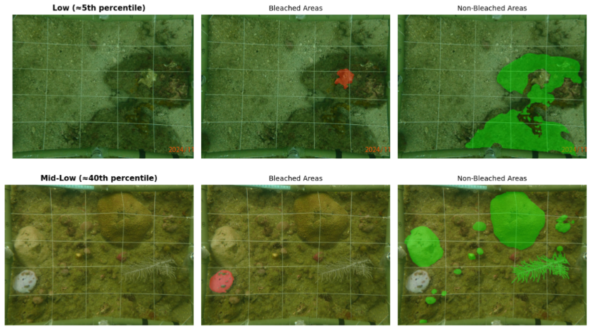

Figure 15: Approach B low-medium example

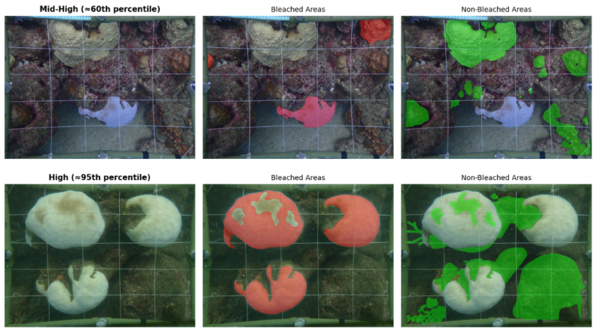

Figure 16: Approach B medium-high example

6.5 Error / Pattern Analysis

Common problems encountered: (i) *spurious whiteness* (sand/ speculars) inflating paleness; mitigated by masking + texture. (ii) *Lighting drift* across sites; mitigated by per-site standardization. (iii) *Small/ fragmented masks* depressing texture, handled by mask QC and flagged in coverage stats.

7 Discussion

Interpretation

Results show two complementary signals: paleness (albedo/luminance/saturation) and micro-texture (Laplacian/GLCM). Both co-vary with the bleaching proxy as expected, indicating the pipeline captures meaningful physiology rather than artefacts.

7.1 Strengths, Weaknesses, Validity

- **Strengths:** transparent features, within-mask computation; site-aware evaluation, dual outputs (index + examples).
- **Weaknesses:** binary labels in data (no ordinal severity), domain shift, residual lighting effects.
- **Validity:** consistency across figures, example galleries provide qualitative triangulation.

7.2 Objective Fulfilment & Stakeholder Impact

Goal of a graded, interpretable assessment using stakeholder data is achieved. The index supports monitoring, exemplars aid trust and review. It has been proven how future work can lead to time and money saving advancements.

7.3 Limitations and Future Work

Add ordinal severity labels, expand sites, and refine color constancy. Explore per-site calibration/test-time adaptation and integrate the index with exemplar explanations in a lightweight UI.

8 Conclusion

A low-compute, explainable pipeline that segments coral, extracts literature-motivated features, and delivers graded health indications was demonstrated. Paleness and texture behave distinctly yet coherently with bleaching, enabling a transparent index and exemplar-based explanations. It is recommend to deploy the feature toolkit for monitoring with human oversight, then iterating toward site-calibrated thresholds and ordinal labels to strengthen reliability.

References

Beijbom, O., Edmunds, P. J., Kline, D. I., Neal, B. P., Mitchell, B. G., & Kriegman, D. (2012). Automated annotation of coral reef survey images [IEEE Xplore document 6247798]. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. https://ieeexplore.ieee.org/document/6247798

CoralNet. (n.d.). *Coralnet*. Retrieved October 26, 2025, from https://coralnet.ucsd.edu/about/

CoralWatch, The University of Queensland. (n.d.). *Using the chart*. Retrieved October 26, 2025, from https://coralwatch.org/monitoring/using-the-chart/

NOAA. (n.d.). *Corals tutorial*. Retrieved October 25, 2025, from https://oceanservice.noaa.gov/education/tutorial_corals/coral07_importance.html

Valinsky, E. (2015, April 14). *Art brings the coral reef crisis above the surface*. Mission Blue. Retrieved October 26, 2025, from https://missionblue.org/2014/05/art-brings-the-coral-reef-crisis-above-the-surface/

Van Woesik, R., Shlesinger, T., Grottoli, A. G., Toonen, R. J., Vega Thurber, R., Warner, M. E., Hulver, A. M., Chapron, L., McLachlan, R. H., Albright, R., Crandall, E., DeCarlo, T. M., Donovan, M. K., Eirin-López, J., Harrison, H. B., Heron, S. F., Huang, D., Humanes, A., Krueger, T., & Zaneveld, J. (2022). Coral-bleaching responses to climate change across biological scales. *Global Change Biology*, *28*(14), 4229–4250. https://doi.org/10.1111/gcb.16192

A Hyperparameter Optimization

Table 3: Search space for segmentation model hyperparameters

Category	**Values / Range**
Encoder/backbone	EfficientNet-B0, {EffNet-B3}, {ResNet34}
Input size	512×512 (fixed)
Batch size	{8, 12, 16}
Learning rate (LR)	$\{1 \times 10^{-3}, 5 \times 10^{-4}, 3 \times 10^{-4}, 1 \times 10^{-4}\}$
Weight decay (WD)	$\{0, 1 \times 10^{-6}, 1 \times 10^{-5}\}$
Loss (BCE:Dice)	{(0.5:0.5), (0.7:0.3), (0.8:0.2)}
Scheduler	ReduceLROnPlateau (patience={5,7}, factor={0.5,0.2})
Epochs (max)	{40, 60} with early stopping (patience={8,10})
Augmentations	flips; rotations ±15°; brightness/contrast $\pm\{0.1, 0.2\}$; saturation $\pm\{0.1, 0.2\}$;
hue shift $\pm\{5°, 10°\}$; mild blur $\sigma \leq \{1.0\}$; CLAHE; coarse dropout $p = \{0.05, 0.1\}$	
Val/Test policy	normalization only (no augmentations)
Split protocol	site-stratified train/val/test (fixed across trials)

Table 4: Top trial results under site-held-out validation. Best model highlighted.

Trial	Encoder	LR	WD	BCE:Dice	Aug set	mIoU	Dice	Notes
T01	EffNet-B0	0.00	0	0.5:0.5	basic	0.66	0.80	baseline
T07	EffNet-B0	0.00	1.00×10^{-5}	0.7:0.3	basic+CLAHE	0.65	0.81	+contrast stability
T12	**EffNet-B0**	0.00	$\mathbf{1.00 \times 10^{-5}}$	**0.7:0.3**	**basic+CLAHE+coarse**	**0.669**	**0.820**	**chosen**
T14	EffNet-B3	0.00	1.00×10^{-5}	0.7:0.3	basic	0.66	0.82	heavier encoder
T18	ResNet34	0.00	1.00×10^{-6}	0.8:0.2	basic	0.66	0.81	alt backbone

Table 5: Final training configuration used for results in the paper.

Item	Value
Encoder/backbone	EfficientNet-B0 (ImageNet init)
Input / Output	$3 \times 512 \times 512$ RGB $\rightarrow 1 \times 512 \times 512$ sigmoid
Loss	BCEWithLogits:Dice = 0.7 : 0.3
Optimizer	AdamW (LR = 3×10^{-4}, WD = 10^{-5})
Scheduler	ReduceLROnPlateau (patience 7, factor 0.5)
Epochs / Early stop	max 60 / patience 10
Batch size	12
Augmentations	flips, $\pm 15°$ rotations, brightness/contrast/saturation ± 0.2, hue $\pm 10°$, mild blur, CLAHE, coarse dropout $p = 0.1$
Val/Test	normalization only
Split	site-stratified (fixed folds)

B Detailed Interpretation of Results Figures

A. Color/Whiteness Correlations

Figure: Figure 8.
Read it. Albedo and luminance move together and align with the composite whiteness/paleness signal; saturation and raw red vary in the opposite direction. The bleaching proxy follows the paleness block.
Why it matters. Confirms pigmentation loss is captured by brightness/whiteness metrics, with saturation/red providing counter-evidence of pigment presence.
Caveats. Illumination can inflate luminance; computing within masks and using albedo mitigates (not eliminates) shading effects.

B. Texture Correlations

Figure: Figure 9.
Read it. Laplacian and LBP share micro-texture information; GLCM correlation trends inversely with high-frequency detail, consistent with smoother, more self-similar structure under bleaching. The bleaching proxy co-varies with increasing regularity and reduced high-frequency content.
Why it matters. Texture captures structural change distinct from color, justifying its inclusion alongside whiteness.

C. Texture Scatter Matrix

Figure: Figure 10.
Read it. As Laplacian/LPB increase, points generally shift toward lower bleaching; GLCM correlation shows the converse. Marginals reveal site/scene effects.
Why it matters. Monotonic but noisy trends indicate complementary texture signals across sites.

D. Whiteness Scatter Matrix

Figure: Figure 11.
Read it. Albedo and luminance rise together; saturation and red fall as paleness increases. Columns against the bleaching proxy show that paler, less saturated tissue is more bleached.
Why it matters. Multiple color-space views agree on the pigment-loss pattern.

E. Texture vs. Bleaching (Scatter)

Figure: Figure 12.
Read it. Bleaching aligns with reduced Laplacian/LPB (loss of fine detail) and increased GLCM correlation (more regular grey-level dependence).
Why it matters. Per-image relationships match the correlation blocks, supporting texture as an independent indicator.

F. Whiteness vs. Bleaching (Scatter)

Figure: Figure 13.
Read it. Brighter/whiter and less saturated tissue corresponds to higher bleaching.
Why it matters. Validates whiteness as a practical severity proxy in the absence of ordinal labels.
Caveats. Sand/specular highlights can spuriously increase paleness; within-mask computation and texture cues reduce such errors.

G. Approach A - Regression Demo

Figure: Figure 14.
Read it. Images sorted by the index progress from darker, textured colonies (higher health) to paler, smoother ones (lower health). Coefficients in Tables 1–2 preserve expected signs.
Why it matters. A transparent linear combination reproduces intuitive health ordering without black-box modeling.
Caveats. Absolute scale depends on per-site calibration (recommended for deployment).

H. Approach B - Exemplar Galleries

Figures: Figure 15, Figure 16.
Read it. Galleries at 5th/40th/60th/95th percentiles shift from textured, pigmented tissue toward paler, smoother colonies as percentiles rise in the paleness/low-texture directions.
Why it matters. Percentile-matched exemplars provide qualitative triangulation and explainability for stakeholders, with minimal compute.
Caveats. Rare morphologies or lighting extremes can bias nearest-percentile matches; per-site caps and feature weighting help.

I. Synthesis and Practical Use

Synthesis. Whiteness (albedo/luminance/saturation/red) and texture (Laplacian/GLCM/LPB) respond in complementary directions with bleaching; neither family alone is sufficient across varied sites.
Use. Report both sub-scores and the composite index; attach exemplar panels for borderline cases. Prefer per-site calibration and human-in-the-loop review for decisions with consequences.