INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 11, November 2026 | ISSN: 2456-4184 | IJNRD.ORG

AI-BASED CORAL REEF HEALTH MONITORING USING UNDERWATER IMAGE PROCESSING

Transfer-Learning Comparison, Ensemble Prediction and Preprocessing Ablation

Dr. Santosh Singh[1], Amit Kumar Pandey[2], Hritikesh Pandey[3], Mohammed Saqlain Shaikh[4]

[1]Head of Department, IT, Thakur College of Science and Commerce, Mumbai 400101, India, sksingh@tcsc.edu.in

[2]Assistant Professor, Thakur College of Science and Commerce, Mumbai 400101, India, amitpandey8089@gmail.com

[3,4] PG Student, Thakur College of Science and Commerce, Mumbai 400101, India, hritikeshpandey18@gmail.com, saqusecond@gmail.com

Abstract: An AI-assisted pipeline can classify underwater coral images as Healthy or Bleached with about 78% cross-validated accuracy, but the evidence is specific to a small dataset and does not yet cover unseen reefs. Manual assessment of large coral photograph collections is slow, subjective and hard to scale, and underwater imaging adds colour distortion, low contrast and turbidity. This paper presents an integrated workflow that combines underwater image enhancement, transfer learning, probability-averaging ensemble prediction, Grad-CAM visual explanation and confidence-based human review. On 896 de-duplicated images (473 bleached, 423 healthy), a from-scratch CNN is compared with four ImageNet-pretrained networks (MobileNetV2, ResNet50, EfficientNetB0, DenseNet121) under identical training conditions, using five-fold stratified cross-validation and a fixed 135-image test set with bootstrap confidence intervals. Pretrained models reach 72.7–75.9% mean accuracy against 57.3% for the baseline, and the four-model ensemble records the highest mean (77.7 ± 4.5% accuracy, 0.775 macro-F1, 0.864 ROC-AUC), although variability among the pretrained models overlaps. A MobileNetV2 ablation shows that a Gray-World, CLAHE, bilateral-filter and unsharp-mask enhancement chain did not improve accuracy (74.2 ± 5.6% raw versus 74.1 ± 3.2% enhanced). Results are reported with their uncertainty, and generalization limits are stated explicitly.

Index Terms - coral bleaching; underwater image processing; deep learning; transfer learning; ensemble learning; explainable AI; Grad-CAM; marine monitoring.

I. INTRODUCTION

Coral reefs shelter fish and invertebrates, buffer shorelines against waves and support fisheries and tourism, yet they face rising sea temperatures, pollution, overfishing, sedimentation and ocean acidification. Thermal stress can make corals expel the symbiotic algae in their tissues, producing the pale appearance known as bleaching; prolonged bleaching can lead to mortality and loss of reef structure. Regular monitoring is therefore a precondition for conservation planning.

Conventional monitoring relies on divers who collect photographs that experts later inspect by eye. This approach is accurate but costly in time, equipment and personnel, depends on weather and visibility, and varies between observers as image collections grow. Deep-learning systems such as the CoralNet classifier engine [1] and recent coral health classifiers [2] show that part of this effort can be automated.

Underwater images, however, are not ordinary photographs. Wavelength-dependent absorption produces blue-green colour casts, scattering lowers contrast, and suspended particles add noise [3]–[5]. A classifier may therefore respond to acquisition conditions rather than coral condition. In addition, a bare class label gives scientists little basis for trusting a prediction, which motivates explainable methods such as Grad-CAM [6].

This paper studies the full pipeline from image to interpretable prediction rather than accuracy alone. Its contributions are: (1) a reproducible binary (Healthy/Bleached) workflow combining underwater enhancement, two-stage transfer learning, ensemble averaging, Grad-CAM explanation and confidence-based review; (2) a controlled comparison of a from-scratch CNN and four ImageNet-pretrained architectures under identical settings, evaluated with five-fold cross-validation, a fixed test set and bootstrap confidence intervals; and (3) a preprocessing ablation showing that, in this experiment, classical enhancement did not improve accuracy, with an explicit account of the uncertainty and limitations of every result.

II. LITERATURE REVIEW

Coral-specific classification has moved from handcrafted features to convolutional networks and transfer learning. CoralNet's deep-learning engine adopted an EfficientNet-based backbone for point annotation [1]; Ghaffar and Choi proposed a two-stream network with Grad-CAM interpretation and class-imbalance handling for health classification [2]; ML-Net uses multi-local perception for healthy versus bleached images [7]; and Menaga et al. compared MobileNetV2, VGG16 and ResNet on coral health [8]. Segmentation studies extend this to spatial analysis, for example U-Net-based survey mapping [10] and the Coralscapes dense-annotation dataset [11].

Underwater enhancement research, including physics-based correction [3], surveys [4] and benchmarks such as UIEB

[5], shows that restoring colour and contrast improves visual quality. Whether such restoration helps downstream classification is a separate empirical question. Cross-domain evidence from ReefNet indicates that models can lose accuracy when tested on sources different from those used in training [9], so random-split results should not be read as field performance. ImageNet pretraining remains useful mainly when target data are limited [12], which suits small coral datasets.

The gap addressed here is not the absence of coral classifiers, but the scarcity of studies that evaluate one integrated workflow (enhancement, several architectures, ensembling, explanation, confidence-aware review) with uncertainty reported and the value of enhancement tested rather than assumed.

III. METHODOLOGY

The system classifies each underwater image as Healthy or Bleached and returns class probabilities, a Grad-CAM heatmap and a flag for low-confidence cases (Fig. 1). It is an image-level classifier: it does not estimate coral cover, species or reef-scale condition.

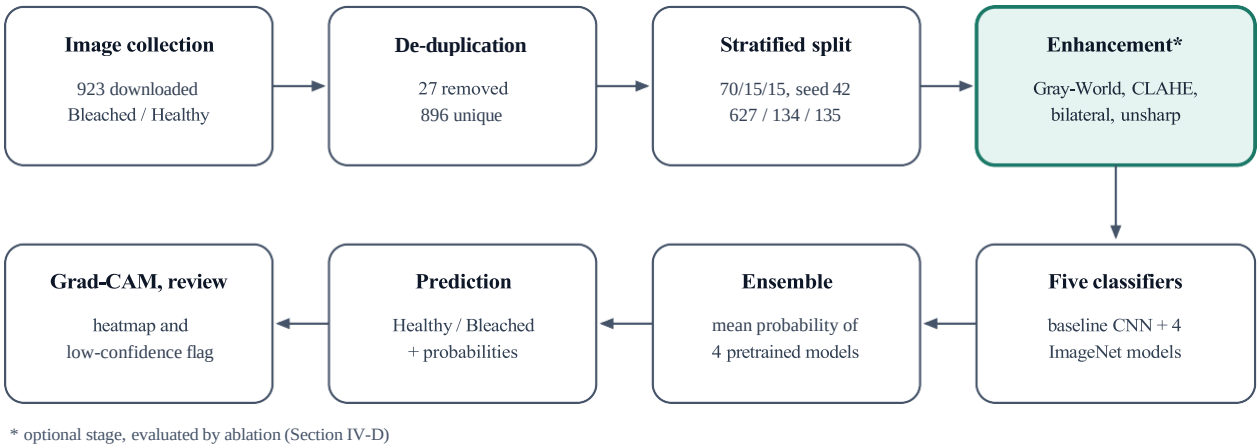


Fig. 1. Processing pipeline. Enhancement (highlighted) is optional and is evaluated by ablation; the baseline CNN is excluded from the ensemble.

A. Dataset

The collection holds 923 downloaded images (485 bleached, 438 healthy). Removing 27 duplicates (12 bleached, 15 healthy) before partitioning leaves 896 unique images (473 bleached, 423 healthy) and prevents identical images from appearing in both training and evaluation data. A stratified 70/15/15 split with seed 42 gives 627 training, 134 validation and 135 test images (71 bleached, 64 healthy in the test set). All inputs are 224 × 224 × 3 RGB. Class weights of 0.947 (bleached) and 1.059 (healthy) offset the moderate imbalance.

B. Underwater enhancement

Four operations are applied in sequence: Gray-World colour correction [13] to reduce the global colour cast; CLAHE [14] with clip limit 2.0 and 8 × 8 tiles for local contrast; bilateral filtering [15] ($d =$, $\sigma = 50$) for edge-preserving denoising; and unsharp masking to restore detail. Whether this branch helps classification is tested by ablation (Section IV-D).

C. Models and training

Five classifiers are compared: a baseline CNN trained from scratch (1.18 M parameters) and four ImageNet-initialized networks, MobileNetV2 [17] (2.26 M), ResNet50 [16] (23.59 M), EfficientNetB0 [18] (4.05 M) and DenseNet121 [19] (7.04 M). Pretrained models use two-stage transfer learning: Stage 1 freezes the backbone and trains the new classification head (learning rate 10^{-3}, at most 15 epochs); Stage 2 unfreezes the top 25% of the backbone and fine-tunes at 10^{-5} (at most 20 epochs). All models share the Adam optimizer [20], batch size 16, sparse categorical cross-entropy with label smoothing 0.05, dropout 0.3, early stopping (patience 6) and ReduceLROnPlateau. Augmentation is applied to training images only: horizontal and vertical flips, rotation (about ±90°), zoom 0.15, translation 0.10, contrast 0.15 and brightness 0.15.

D. Ensemble, explanation and review

The ensemble averages the class probabilities of the four pretrained networks (the baseline is excluded); the class with the larger mean probability is the prediction:

$$p_{ens}(c \mid x) = \frac{1}{4} \sum_i p_i(c \mid x) \quad (1)$$

where *i* runs over MobileNetV2, ResNet50, EfficientNetB0 and DenseNet121. Grad-CAM [6] overlays on the input the regions that contributed most to a prediction, and low-confidence predictions are flagged for a human reviewer. These two components are provided by the implemented system; this paper presents them as design features and does not evaluate them quantitatively.

E. Evaluation protocol

Five-fold stratified cross-validation reports mean ± standard deviation of accuracy, precision, recall, macro-F1 and ROC-AUC. The fixed 135-image test set reports accuracy with 95% bootstrap confidence intervals, macro-F1 and ROC-AUC. Training time and per-image inference time are also recorded.

IV. RESULTS

A. Cross-validation

Pretrained models clearly outperform the from-scratch baseline, and the ensemble has the highest mean on every reported metric (Table I, Fig. 2). The baseline reaches only 57.3 ± 9.8% accuracy with unstable precision (0.370 ± 0.237), whereas the four pretrained networks lie between 72.7% and 75.9%. ResNet50 has the highest single-model mean (75.9 ± 2.3%) and the lowest fold-to-fold variation. The ensemble records 77.7 ± 4.5% accuracy, 0.775 macro-F1 and 0.864 ROC-AUC. Standard deviations of the four pretrained models overlap, so their ranking is not statistically separated.

TABLE I. Five-fold stratified cross-validation (mean ± SD)

Model	Acc.	Prec.	Recall	Macro-F1	ROC-AUC
Baseline CNN	0.573 ± 0.098	0.370 ± 0.237	0.548 ± 0.106	0.423 ± 0.173	0.666 ± 0.168
MobileNetV2	0.741 ± 0.032	0.748 ± 0.041	0.743 ± 0.036	0.740 ± 0.032	0.834 ± 0.029
ResNet50	0.759 ± 0.023	0.763 ± 0.022	0.757 ± 0.022	0.756 ± 0.023	0.858 ± 0.043
EfficientNetB0	0.751 ± 0.049	0.763 ± 0.047	0.753 ± 0.040	0.748 ± 0.050	0.825 ± 0.049
DenseNet121	0.727 ± 0.051	0.738 ± 0.053	0.728 ± 0.057	0.723 ± 0.058	0.808 ± 0.058
Ensemble (4)	0.777 ± 0.045	—	—	0.775 ± 0.047	0.864 ± 0.043

Ensemble precision and recall were not recorded (—).

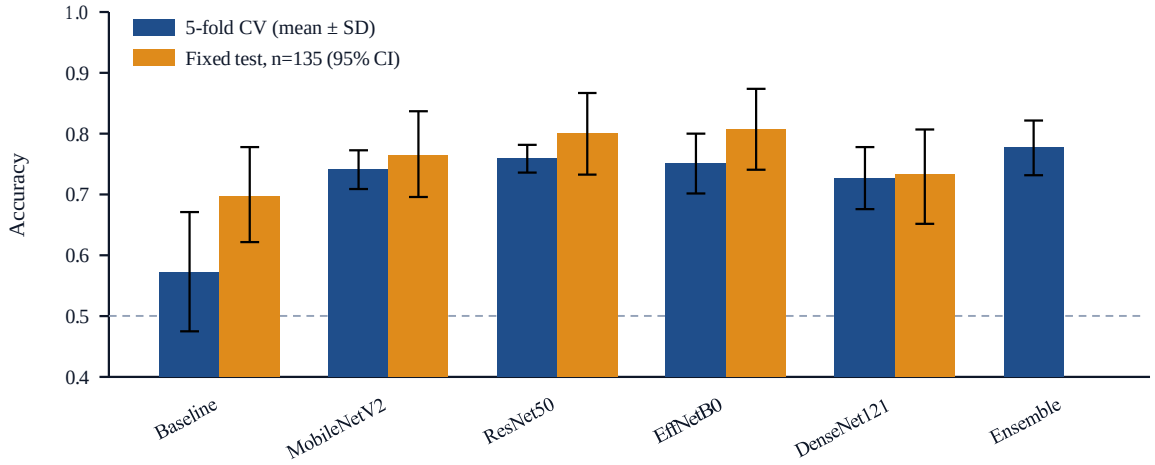


Fig. 2. Accuracy under both protocols (error bars: ±1 SD for cross-validation, 95% bootstrap CI for the fixed test). The y-axis starts at 0.4; the dotted line marks 0.5.

B. Fixed test set

On the 135 held-out images, EfficientNetB0 (0.807) and ResNet50 (0.800) score highest (Table II), and ResNet50 has the best ROC-AUC (0.914). The 95% bootstrap intervals are wide, about ±7–8 percentage points, and overlap across the four pretrained models, so differences of a few points are not decisive. Fixed-test and cross-validation values differ because a single test set is sensitive to which images it contains.

TABLE II. Fixed test-set performance (n = 135)

Model	Accuracy [95% CI]	Macro-F1	ROC-AUC
Baseline CNN	0.696 [0.622, 0.778]	0.686	0.807
MobileNetV2	0.763 [0.696, 0.837]	0.762	0.870
ResNet50	0.800 [0.733, 0.867]	0.800	0.914
EfficientNetB0	0.807 [0.741, 0.874]	0.807	0.870
DenseNet121	0.733 [0.652, 0.807]	0.733	0.800

C. Computational cost

Parameter count does not predict cost (Table III). EfficientNetB0 has 4.05 M parameters but the longest training time (31.91 min), while DenseNet121 has the slowest inference (30.5 ms per image). The baseline is fastest (5.9 ms) but least accurate; MobileNetV2, ResNet50 and EfficientNetB0 run at 15.5–17.4 ms per image. Timings depend on the hardware used and should be re-measured on any deployment target.

TABLE III. Model size and computational cost

Model	Params (M)	Epochs run	Train (min)	Infer. (ms/img)
Baseline CNN	1.18	15	3.91	5.9
MobileNetV2	2.26	21	11.25	15.7
ResNet50	23.59	17	7.20	15.5
EfficientNetB0	4.05	30	31.91	17.4
DenseNet121	7.04	15	6.89	30.5

D. Preprocessing ablation

With MobileNetV2, the enhancement chain did not improve accuracy. On the fixed split, accuracy fell from 0.800 on raw images to 0.763 on enhanced images, a difference of about five of 135 images with overlapping uncertainty. In five-fold cross-validation the means were practically identical: 0.742 ± 0.056 (raw) versus 0.741 ± 0.032 (enhanced). Enhanced inputs showed a smaller standard deviation, but the experiment does not show this is statistically significant, and the mechanism behind the absence of gain was not investigated.

V. DISCUSSION AND LIMITATIONS

Transfer learning is the clearest finding. With fewer than 750 training images, every pretrained network beat the from-scratch baseline by 15 to 19 accuracy points in cross-validation, consistent with the value of pretrained representations when target data are scarce [12]. Averaging the four pretrained models gave the best mean result, plausibly because their errors are partly complementary, but the gain over ResNet50 (1.8 points) is smaller than the fold-to-fold standard deviation, so it should be treated as suggestive.

The ablation result is a useful negative finding. Colour correction, contrast equalization, denoising and sharpening make images look better to humans, yet they did not raise accuracy for MobileNetV2. Fine-tuned networks with training-time augmentation may already absorb much of the colour and contrast variation, and fixed enhancement parameters may remove cues that separate bleached from healthy coral. Preprocessing should therefore be validated against the task instead of assumed beneficial; other models and learned enhancement methods remain untested.

The main limitations are: (i) data — 896 images cover limited species, depths, cameras and sites, and the 135-image test set yields wide confidence intervals; (ii) task definition — the binary label ignores partial bleaching, disease, damage and mortality, so the output is not a reef-scale health score; (iii) generalization — no geographically independent test set was used, and cross-domain studies show random-split accuracy can overstate field performance [9]; regional coverage, for example of Indian reefs, is not claimed; (iv) explainability and confidence — Grad-CAM maps and confidence flags were not quantitatively validated and calibration was not assessed; and (v) ablation scope — enhancement was tested with one architecture only, and ensemble precision and recall were not recorded.

The system is therefore best used as a screening aid that prioritizes images for expert review, not as a replacement for field surveys or marine biologists.

VI. CONCLUSION AND FUTURE WORK

An integrated pipeline of transfer learning, ensemble averaging and Grad-CAM explanation separated healthy from bleached coral images with 77.7 ± 4.5% cross-validated accuracy and 0.864 ROC-AUC, well above a from-scratch CNN at 57.3%. Individual pretrained architectures were statistically indistinguishable, and the tested enhancement chain gave no

accuracy benefit, which argues for evidence-based preprocessing choices.

Future work should (i) enlarge and diversify the dataset and add region-held-out and cross-camera evaluation; (ii) extend labels to multiple health categories and add segmentation for percentage-cover estimates; (iii) test learned enhancement methods across several architectures; (iv) calibrate confidence and evaluate the Grad-CAM maps and human-review workflow quantitatively; and (v) profile deployment on edge devices, underwater robots and video streams.

REFERENCES

1. Q. Chen, O. Beijbom, S. Chan, J. Bouwmeester, and D. Kriegman, "A new deep learning engine for CoralNet," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 3693–3702.
2. A. A. Ghaffar and G. S. Choi, "A two-stream deep learning framework for robust coral reef health classification: Insights and interpretability," *IEEE Access*, vol. 13, pp. 78490–78512, 2025, doi: 10.1109/ACCESS.2025.3561226.
3. D. Akkaynak and T. Treibitz, "Sea-thru: A method for removing water from underwater images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1682–1691.
4. S. Anwar and C. Li, "Diving deeper into underwater image enhancement: A survey," *Signal Process. Image Commun.*, vol. 89, Art. no. 115978, 2020.
5. C. Li et al., "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
6. R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 618–626.
7. S. Wang et al., "ML-Net: A multi-local perception network for healthy and bleached coral image classification," *J. Mar. Sci. Eng.*, vol. 12, no. 8, Art. no. 1266, 2024.
8. D. Menaga, S. Lemi Deborah, and M. Makizhna, "Deep learning models for coral reef health classification: MobileNetV2, VGG16 and ResNet," Atlantis Press, 2025, doi: 10.2991/978-94-6463-718-2_103.
9. Y. Battach et al., "ReefNet: A large-scale, taxonomically enriched dataset and benchmark for hard coral classification," arXiv:2510.16822, 2025.
10. K. Mizuno et al., "An efficient coral survey method based on a large-scale 3-D structure model obtained by Speedy Sea Scanner and U-Net segmentation," *Sci. Rep.*, vol. 10, Art. no. 12416, 2020.
11. J. Sauder et al., "The Coralscapes dataset: Semantic scene understanding in coral reefs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2025, pp. 2136–2143.
12. K. He, R. Girshick, and P. Dollár, "Rethinking ImageNet pre-training," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 4918–4927.
13. G. Buchsbaum, "A spatial processor model for object colour perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, 1980.
14. K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474–485.
15. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 1998, pp. 839–846.
16. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
17. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 4510–4520.
18. M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, PMLR vol. 97, 2019, pp. 6105–6114.
19. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4700–4708.
20. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015.

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.