Coral classification in underwater images using a dual-branch deep learning framework

Pracharat Sa-ngadsup, Chawan Koopipat
Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

Article Info

Article history:

Received Mar 9, 2026
Revised Jun 3, 2026
Accepted Jun 27, 2026

Keywords:

CIELAB color space
Convolutional neural networks
Coral image classification
Coral reef monitoring
Discrete wavelet transform
Feature fusion
Underwater image analysis

ABSTRACT

Automated coral classification from underwater imagery is essential for large-scale reef monitoring but remains challenging due to color attenuation, illumination variability, and differences between texture-focused close-range images and morphology-focused colony-level observations. To address these challenges, this study proposes a dual-branch convolutional neural network that integrates information from the CIELAB (LAB) color space with structural descriptors derived from the discrete wavelet transform (DWT). RGB images are first converted to the LAB color space to separate luminance and chromatic components in separate channels, enabling the model to exploit color and lightness information more explicitly. Structural information is extracted from the luminance channel using wavelet decomposition to capture high-frequency morphological patterns. The two representations are processed through parallel convolutional branches and fused at the feature level for classification. Experiments conducted on a unified coral dataset containing texture-dominant and morphology-focused imagery across 14 classes show that the proposed method achieves 96.52% accuracy on the texture-focused RSMAS dataset and 88.33% accuracy on the morphology-focused structure RSMAS dataset, reducing the cross-domain performance gap from 22.46% to 8.19% compared with RGB baselines, demonstrating improved cross-domain robustness for coral classification under heterogeneous underwater imaging conditions.

Corresponding Author:

Pracharat Sa-ngadsup
Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University
Bangkok 10330, Thailand
Email: pracharat.sa@chula.ac.th

1. INTRODUCTION

Coral reefs are among the most biologically diverse marine ecosystems and provide essential ecological and economic services, including coastal protection, fisheries support, and biodiversity preservation [1]. However, coral reefs worldwide are increasingly threatened by climate change, ocean warming, and anthropogenic disturbances, which accelerate coral bleaching and ecosystem degradation [2]-[4]. Traditional field surveys are labor-intensive and spatially limited, making automated image-based analysis an increasingly important tool for reef monitoring at scale [5].

Early studies on coral image classification primarily relied on handcrafted feature extraction combined with classical machine learning methods such as support vector machines (SVMs) and texture-based descriptors [6]-[9]. While these approaches demonstrated promising results, their performance is often constrained by the limited representational capacity of manually designed features. Classical machine learning methods have also been explored in more recent work [10]. Recent advances in deep learning, particularly convolutional neural networks (CNNs), have significantly improved performance in underwater image classification tasks by enabling automatic hierarchical feature learning [11]-[17].

Despite these advances, coral image classification in underwater environments remains challenging. Underwater images often suffer from color distortion, reduced contrast, and scattering caused by light absorption and suspended particles [18]. Several methods have been proposed to address these issues, including physics-based restoration techniques and learning-based color correction models such as WaterGAN [19]. In addition, underwater image enhancement techniques have been developed to restore color and contrast in degraded marine imagery [20].

Another challenge arises from variations in spatial scale across coral image datasets. Close-range imagery typically emphasizes fine surface textures, whereas colony-level imagery highlights larger morphological structures. When discriminative cues shift from texture-dominated patterns to broader structural characteristics, CNN models trained primarily on texture-rich imagery may exhibit reduced generalization performance. To address this limitation, alternative representation strategies have been explored. Transforming images into the CIELAB (LAB) color space represents luminance and chromatic components in separate channels, which can reduce sensitivity to illumination variability [18], [21]. In parallel, frequency-domain analysis using the discrete wavelet transform (DWT) enables multi-resolution characterization of structural patterns [22]-[24].

Although multi-branch and feature-fusion architectures have demonstrated improved performance in general image classification by processing complementary representations in parallel [25], [26], these designs are typically evaluated in natural-image recognition settings and do not explicitly address the physical and domain-specific characteristics of underwater coral imagery, including wavelength-dependent color attenuation, illumination variability, and changes in visual scale. Existing coral classification methods predominantly rely on single-input RGB representations [11]-[17]. In RGB images, changes in illumination, shading, water attenuation, and camera exposure directly affect the red, green, and blue intensity channels, making brightness and color variation difficult to separate. By contrast, CIELAB represents lightness in the L^* channel and chromaticity along the opponent-color axes a^* and b^*. In principle, regions with similar chromaticity can remain more comparable under moderate lighting differences when luminance is represented separately. This property is useful for coral classification because images acquired at different scales and under different underwater conditions may vary strongly in brightness while retaining species-related color characteristics.

However, color representation alone does not capture colony-level morphology. Color-correction methods such as WaterGAN [19] can improve visual appearance, but they do not directly encode morphology-related structural descriptors for taxonomic discrimination.

Although Gómez-Ríos *et al.* [15] reported results on both texture-focused and morphology-focused coral datasets, their evaluation used a two-level classifier under different experimental conditions. This leaves open the question of whether a single representation framework can maintain performance when the visual cues shift from close-range texture to colony-level morphology. Prior work has not systematically isolated the individual and combined effects of LAB color representation, wavelet-derived structural descriptors, and dual-branch feature fusion on performance across coral image domains.

In this study, the term domain gap refers to the performance difference observed when a model is evaluated across coral image domains with different visual characteristics, particularly texture-focused close-range imagery and morphology-focused colony-level imagery. This gap is interpreted as an indicator of how strongly a model depends on cues that may not transfer across imaging scale, viewpoint, illumination, and structural context. A smaller domain gap therefore suggests better cross-domain generalization, especially when performance on the morphology-focused dataset improves without substantially reducing accuracy on the texture-focused dataset.

To address this gap, the present study proposes a dual-branch deep learning framework in which the LAB branch represents luminance and chromatic components in separate channels, while the DWT branch extracts multi-resolution structural detail from the luminance channel. The two representations are processed through parallel convolutional branches and fused before classification. The framework is evaluated by directly comparing RGB, LAB, DWT-only, RGB+DWT, and LAB+DWT configurations under a unified evaluation protocol.

The main contributions of this study are summarized as follows:
- A dual-branch representation learning framework that combines LAB color representation with DWT-derived structural descriptors;
- A systematic evaluation of RGB, LAB, DWT-only, RGB+DWT, and LAB+DWT configurations across texture-dominant and morphology-focused coral image datasets; and

– A domain-gap analysis that quantifies cross-domain generalization as the absolute difference in classification accuracy between the texture-focused RSMAS dataset and the morphology-focused structure RSMAS dataset.

2. PROPOSED METHOD

The proposed framework integrates LAB color representation with DWT-derived descriptors for coral taxonomy. The overall workflow consists of three main stages: representation transformation, dual-branch feature extraction, and feature fusion for classification, as illustrated in Figure 1.

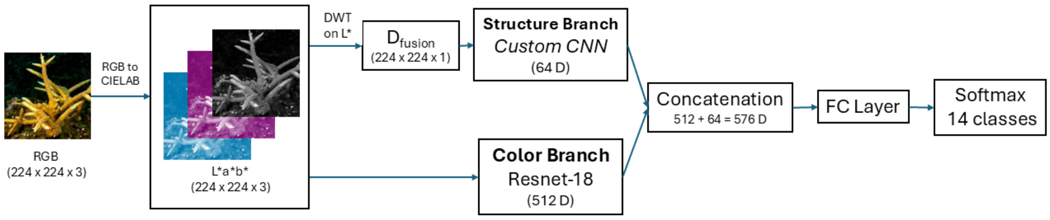


Figure 1. Overview of the dual-branch architecture

2.1. Representation transformation

Underwater images often exhibit significant color distortion due to wavelength-dependent light absorption and scattering. To represent luminance and chromatic components more explicitly, input RGB images are transformed into the CIELAB (LAB) color space, which represents luminance information ($L*$) and chromatic components ($a*$ and $b*$).

In addition to color-space transformation, structural information is extracted using the DWT. Wavelet decomposition performs multi-resolution analysis by separating image content into frequency sub-bands, where low-frequency components represent coarse image structure while high-frequency components capture edges and fine spatial details [22]-[24], [27]-[29]. Because coral colonies exhibit complex surface geometry and branching morphology, these high-frequency components highlight patterns associated with coral growth forms. Consequently, the wavelet representation provides a compact descriptor that complements the LAB color input.

Specifically, the Daubechies-2 (db2) mother wavelet [30] is applied at a single decomposition level to the luminance channel. The db2 wavelet was selected for its balance between compact support (4 coefficients) and smoothness, which is suitable for capturing edge and texture patterns in coral surface morphology. Single-level decomposition preserves spatial resolution while extracting directional high-frequency information.

Let I_l denote the normalized luminance channel extracted from the CIELAB representation. A single-level two-dimensional DWT yields four sub-band coefficient matrices.

$$\{LL, LH, HL, HH\} = DWT_2(I_L; db2) \tag{1}$$

Where LL is the approximation sub-band, and LH, HL, and HH are the horizontal, vertical, and diagonal detail sub-bands, respectively. The approximation sub-band LL is discarded because it primarily contains smoothed luminance information already represented in the LAB color branch. Each detail sub-band is converted to magnitude form to remove directional sign dependency.

$$D_{LH} = |LH|, D_{HL} = |HL|, D_{HH} = |HH| \tag{2}$$

$$D_{fusion} = \alpha(D_{LH} + D_{HL} + D_{HH}) \tag{3}$$

Where α = 0.5 is a scaling constant that balances the combined magnitude of the three detail sub-bands. The DWT-derived structural map, D_{fusion}, is kept in its computed magnitude scale and standardized implicitly by the batch-normalization layers in the DWT branch. Because DWT decomposition reduces spatial dimensions, the fused structural map D_{fusion} is resized to 224×224 pixels using bilinear interpolation to produce $D_{fusion,224}$. The preprocessing pipeline therefore produces two synchronized inputs: a three-channel LAB image $X_{lab} \in \mathbb{R}^{224\times224\times3}$ for the color branch, and a single-channel resized DWT-derived structural map $D_{fusion,224} \in \mathbb{R}^{224\times224\times1}$ for the DWT branch.

2.2. Dual-branch feature extraction

The proposed architecture consists of two parallel branches that process complementary representations of the input image. The first branch receives the LAB color representation and extracts color-related features using a pretrained ResNet-18 backbone [31], producing a 512-dimensional feature vector from the global average pooling layer. This branch learns hierarchical visual features from the three-channel LAB representation, where luminance and chromatic components are provided as separate input channels.

The second branch processes the DWT-derived structural map extracted from the luminance channel. Because the wavelet representation emphasizes edges and local structural patterns related to coral morphology, it provides an input for extracting shape-related features. A lightweight custom CNN is used for this branch, consisting of three convolutional blocks each comprising a 3×3 convolution, batch normalization, and ReLU activation, with filter sizes of 16, 32, and 64, respectively. A global average pooling layer then produces a 64-dimensional feature vector.

The selection of ResNet-18 [31] as the backbone for the color branch is motivated by several considerations. First, the relatively small coral dataset (1,175 total images) favors a lightweight architecture to reduce the risk of overfitting. Compared with deeper variants such as ResNet-50, ResNet-18 provides sufficient representational capacity while requiring fewer parameters and lower computational cost, making it suitable for the proposed dual-branch framework.

The DWT branch employs a shallow three-layer CNN rather than a deeper pretrained backbone because the DWT-derived structural map is a single-channel representation that differs fundamentally from natural RGB images, making ImageNet-pretrained weights unsuitable for this input modality. The progressive filter expansion from 16 to 64 captures increasingly complex local patterns, while global average pooling produces a compact 64-dimensional feature vector. This asymmetric design, with a pretrained backbone for the LAB branch and a shallow custom network for the DWT branch, reflects the different complexity requirements of each representation.

2.3. Feature fusion and classification

Feature fusion combines color-related and DWT-derived features by concatenating the 512-dimensional LAB branch output with the 64-dimensional DWT branch output, yielding a 576-dimensional fused vector. This vector is passed through a batch normalization layer, a fully connected layer with 256 units followed by ReLU activation and dropout ($p = 0.3$), and a final fully connected layer with 14 output units corresponding to the coral classes. A SoftMax function produces the final class probability distribution.

Feature-level concatenation was chosen over alternative fusion strategies, such as element-wise addition or attention-based fusion, because it preserves the full output of both branches without assuming direct feature alignment. The batch normalization layer applied after concatenation normalizes the feature scales from the two branches, and dropout ($p = 0.3$) reduces co-adaptation between branch features during training.

3. METHOD

3.1. Dataset description

Experimental evaluation was conducted using a unified dataset constructed from two complementary coral image collections: the RSMAS dataset and the structure RSMAS dataset. The RSMAS dataset contains close-range coral imagery emphasizing fine surface textures, while the structure RSMAS dataset focuses on colony-level morphology captured from greater viewing distances and more complex scene contexts [15]. A summary of the dataset composition is presented in Table 1 [4], [15].

Table 1. Dataset summary

Dataset	Classes	Images	Image type
RSMAS [4]	14	766	Texture-focused
Structure RSMAS [15]	14	409	Morphology-focused

All images were collected under uncontrolled underwater conditions and exhibit substantial variation in illumination, color attenuation, scale, and background clutter. Example images from both datasets are presented in Figure 2.

Figure 2. Example coral images from the RSMAS dataset (texture-focused imagery) and the structure RSMAS dataset (morphology-focused imagery)

3.2. Training configuration

All model training and classification experiments were implemented using MATLAB R2025a with the deep learning toolbox and image processing toolbox on a workstation equipped with an NVIDIA GPU. Input images were resized to 224×224 pixels, and the preprocessing pipeline generated two parallel inputs: a three-channel LAB color image for the ResNet-18 branch and a single-channel DWT-derived structural map for the DWT branch. The training configuration is summarized in Table 2. Inference-time benchmarking was conducted separately under CPU-only execution, as described in Section 4.3.

Model training used the Adam optimizer with an initial learning rate of 1×10^{-4} and a mini-batch size of 32. The network was trained for 30 epochs, and the model checkpoint with the lowest validation loss was selected for evaluation (best-validation-loss recovery). Data augmentation included random rotation (±180°), random horizontal reflection, and random vertical reflection, applied only during training. No color jittering or brightness augmentation was applied, as the LAB transformation represents luminance and chromatic components in separate channels.

Table 2. Training parameters

Parameter	Value
Backbone	ResNet-18
Optimizer	Adam
Learning rate	1×10^{-4}
Epoch	30
Batch size	32
Input size	224×224

3.3. Cross-validation protocol

To evaluate model stability and reduce dependence on a single data partition, five-fold cross-validation was performed. In each fold, the dataset was divided into training and validation subsets while maintaining class balance across the partitions.

For each fold, a model was trained independently and evaluated on the corresponding validation subset. Final performance metrics were obtained by averaging the results across the five folds. This procedure provides a more reliable estimate of model performance under varying data splits.

3.4. Evaluation metrics

Classification performance was evaluated using accuracy and macro F1-score. Accuracy measures the proportion of correctly classified samples across all classes, providing an overall measure of classification performance. Macro F1-score provides a balanced evaluation across classes by computing the F1-score independently for each class and then averaging the results. This metric is particularly useful for multi-class classification tasks because it treats all classes equally regardless of class frequency. Accuracy is defined as:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

where TP, TN, FP, and FN represent the number of true positives, true negatives, false positives, and false negatives, respectively. For multi-class classification, accuracy is computed in a one-vs-rest manner, where these values are aggregated across all classes.

The F1-score combines precision and recall into a single harmonic mean measure for each class. In this study, the macro F1-score is computed by averaging the F1-scores across all classes:

$$Macro\ F1 = \left(\frac{1}{K}\right)\sum_{k=1}^{K} F1_k \quad (5)$$

where K represents the number of classes and $F1_k$ denotes the F1-score for class k. For each class,

$$F1_k = \frac{2P_kR_k}{P_k+R_k} \quad (6)$$

where P_k and R_k denote precision and recall for class k, respectively.

3.5. Confusion matrix evaluation

To analyze class-level classification behavior, confusion matrices were generated separately for each dataset domain. Row-wise normalization was applied so that each row represents the proportion of predictions relative to the true class. This evaluation enables identification of coral categories that are difficult to distinguish and provides insight into misclassification patterns observed in both texture-dominant and morphology-focused imagery.

4. RESULTS AND DISCUSSION

Table 3 summarizes the classification performance of baseline and proposed models across both coral image datasets. Conventional RGB models achieved high accuracy under texture-rich RSMAS conditions but showed lower accuracy on morphology-focused structure RSMAS images. In contrast, the proposed LAB+DWT model improved structure RSMAS accuracy and reduced the dataset-level domain gap by combining LAB-based color representation with wavelet-derived structural descriptors. For quantitative comparison, the domain gap was computed as the absolute difference in classification accuracy between the two datasets. This comparison shows that mixed-domain accuracy and domain-gap reduction are related but distinct evaluation objectives.

Table 3. Main comparative and ablation performance of baseline and proposed models under the single best-fold evaluation protocol

Model architecture	Color	Structure	RSMAS accuracy (%)	RSMAS F1	Structure RSMAS accuracy (%)	Structure F1	Domain gap (%)	CV accuracy (%)	CV F1
DWT-only	—	DWT(L)	20.00	0.138	15.00	0.109	5.00	17.03±1.92	0.130
RGB (ResNet-18)	RGB	—	98.26	0.982	83.33	0.823	14.93	92.85±1.52	0.927
RGB (ResNet-50)	RGB	—	99.13	0.984	76.67	0.763	22.46	91.66±2.86	0.916
RGB+DWT	RGB	DWT(L)	94.78	0.937	76.67	0.752	18.11	93.05±2.21	0.927
LAB (ResNet-18)	L*a*b*	—	98.26	0.982	86.67	0.882	11.59	90.29±2.23	0.902
Proposed LAB+DWT	L*a*b*	DWT(L)	96.52	0.974	88.33	0.888	8.19	89.40±2.08	0.889

Note: Test-set accuracy, F1-score, and domain gap are reported under the single best-fold evaluation protocol; CV Accuracy and CV F1 summarize 5-fold cross-validation performance.

4.1. Performance, ablation study, and complementarity of LAB and DWT

The ablation configurations in Table 3 isolate the contribution of each component. The DWT-only configuration achieves only 17.03% cross-validation accuracy, indicating that the structural descriptors provide weak standalone discriminative power and that the DWT branch functions as a complementary feature source rather than a self-sufficient classifier.

The RGB+DWT configuration achieves the highest mixed-domain cross-validation accuracy among the evaluated configurations, but it does not reduce the domain gap. Its domain gap of 18.11% is larger than that of the RGB ResNet-18 baseline (14.93%), suggesting that structural descriptors paired with RGB do not sufficiently address cross-domain variation. In contrast, the proposed LAB+DWT achieves the smallest domain gap among the color-based classifiers (8.19%) and the highest structure RSMAS accuracy (88.33%), although it does not maximize mixed-domain cross-validation accuracy.

These results indicate that DWT is not universally beneficial by itself. LAB represents luminance and chromaticity in separate channels, while DWT captures high-frequency structural cues from the luminance channel. The ablation shows that DWT contributes most effectively when paired with the LAB representation, suggesting that structural descriptors and luminance/chromatic separation act together to reduce the domain gap across imaging conditions.

4.2. Ensemble evaluation and domain-gap analysis

Table 4 reports the five-fold ensemble accuracy, ensemble gain, and ensemble domain gap for each configuration. Ensemble averaging improved structure RSMAS accuracy across most configurations, most substantially for RGB ResNet-50 (+10.00%), followed by RGB+DWT (+5.00%). For the proposed LAB+DWT, ensemble gain on structure RSMAS was modest (+1.67%), reflecting that the single best-fold model already achieves high structure RSMAS accuracy (88.33%).

Among the color-based classifiers, the proposed LAB+DWT achieves the smallest ensemble domain gap (9.13%) and the highest structure RSMAS ensemble accuracy (90.00%). The comparison with RGB+DWT is particularly informative: RGB+DWT achieves higher mixed-domain cross-validation accuracy (93.05% vs. 89.40%) but a larger ensemble domain gap (16.59% vs. 9.13%), indicating that mixed-domain accuracy and cross-domain robustness reflect distinct aspects of model behavior.

Table 4. Ensemble evaluation and domain-gap analysis across evaluated configurations

Model	RSMAS ensemble (%)	RSMAS single gain (%)	Structure ensemble (%)	Structure gain (%)	Ensemble gap (%)
DWT-only	21.74	+1.74	18.33	+3.33	3.41
RGB (ResNet-18)	99.13	+0.87	83.33	0.00	15.80
RGB (ResNet-50)	99.13	0.00	86.67	+10.00	12.46
RGB+DWT	98.26	+3.48	81.67	+5.00	16.59
LAB (ResNet-18)	98.26	0.00	86.67	0.00	11.59
Proposed LAB+DWT	99.13	+2.61	90.00	+1.67	9.13

Note: Ensemble Gap = |RSMAS Ensemble (%) − Structure Ensemble (%)|. Gain = ensemble accuracy − single best-fold accuracy (from Table 3). DWT-only is included for ablation completeness but is not considered a practical classifier.

4.3. Comparison with prior methods

The proposed LAB+DWT framework addresses the cross-domain gap directly. Compared with Lumini *et al.* [8], which reported 99.20% accuracy on RSMAS, the proposed LAB+DWT model achieves a slightly lower RSMAS accuracy of 96.52%. However, the proposed model also achieves 88.33% accuracy on structure RSMAS (Table 5), which was not evaluated by Lumini *et al.* [8], and reduces the domain gap to 8.19%. This trade-off indicates that the proposed dual-branch representation accepts a small reduction in top-line RSMAS accuracy while improving generalization to morphology-focused imagery under a unified evaluation protocol.

Table 5. Comparison with prior methods on RSMAS and structure RSMAS datasets

Method	Input	RSMAS accuracy (%)	Structure RSMAS accuracy (%)	Domain gap (%)
Shihavuddin *et al.* (2013) [9]	RGB	92.74%	—	—
Gómez-Ríos *et al.* (2019) [15]	RGB	98.36%	85.26%	13.10%
Lumini *et al.* (2023) [8]	RGB (CNN ensemble)	99.20%	—	—
Proposed LAB+DWT	L*a*b*+DWT	96.52%	88.33%	8.19%
Proposed LAB+DWT + Ensemble	L*a*b*+DWT	99.13%	90.00%	9.13%

Note: Methods not evaluated on a dataset are marked as "—". Domain gap is computed as the absolute difference in accuracy between the two datasets. The ensemble row reports the five-fold ensemble result for both RSMAS and Structure RSMAS.

4.4. Statistical significance

Table 6 reports paired t-tests across the 5-fold cross-validation accuracy to assess whether the proposed LAB+DWT model significantly improves mixed-domain validation performance compared with the baseline and ablation configurations. The comparison with DWT-only is highly significant ($p < 0.001$), indicating that the DWT structural branch is not sufficient as a standalone classifier and requires the color branch. No statistically significant difference was detected among the comparisons between LAB+DWT and the color-based configurations at $\alpha = 0.05$. Because the main objective of this study is reducing the domain gap rather than maximizing mixed-domain validation accuracy, the paired t-test results should be interpreted together with the ensemble domain-gap analysis presented in Table 4.

Table 6. Statistical significance of cross-validation accuracy differences (paired t-test, 5-fold CV)

Comparison	Mean difference (%)	t-stat	p-value	95% CI
LAB+DWT vs. DWT-only	+72.37	63.359	< 0.001	[+69.20, +75.54]
LAB+DWT vs. RGB (ResNet-50)	−2.26	−1.631	0.178	[−6.11, +1.59]
LAB+DWT vs. RGB (ResNet-18)	−3.45	−2.415	0.073	[−7.42, +0.52]
LAB+DWT vs. LAB (ResNet-18)	−0.89	−2.087	0.105	[−2.08, +0.30]
LAB+DWT vs. RGB+DWT	−3.65	−1.971	0.120	[−8.79, +1.49]

4.5. Computational efficiency

Table 7 summarizes the computational complexity and inference time of the evaluated model configurations. Timing was measured using the selected single-fold model on 175 test images (115 RSMAS and 60 structure RSMAS) and includes preprocessing, LAB conversion, DWT computation, and forward propagation. Timing experiments were conducted in MATLAB R2025a on a Windows 11 Pro desktop with an AMD Ryzen 5 5600 6-core/12-thread CPU at 3.50 GHz and 32 GB RAM using CPU execution only. The proposed dual-branch model adds only 0.18M parameters (+1.6%) over the single-branch ResNet-18 baseline, consisting of the lightweight DWT branch and fusion head. The estimated model size of 43.4 MB is consistent with the single-branch ResNet-18 baseline (42.8 MB), confirming that the dual-branch extension introduces negligible storage overhead. Training time for the proposed LAB+DWT (44.7 min/fold) is 56% of the RGB ResNet-50 training time (80.2 min/fold) while achieving a smaller ensemble domain gap. Per-image inference time is comparable across models under CPU execution, with the proposed LAB+DWT requiring 0.031 seconds per image. For batch processing of underwater survey imagery, where real-time performance is not required, the computational overhead is limited.

Table 7. Model complexity and inference-time comparison

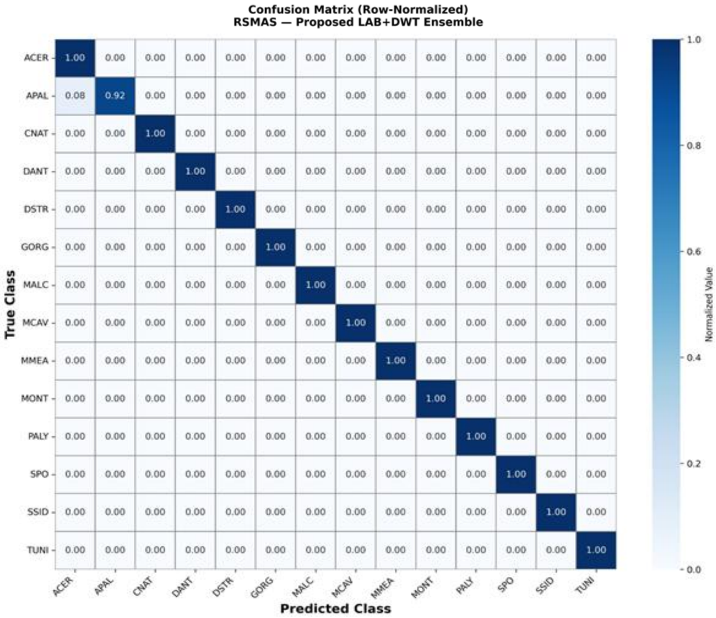
Model	Input	Parameters (M)	Size (MB)	Training time (min/fold)	Inference time Total (s)	Inference time per image (s)
DWT-only	DWT(L)	0.18	0.7	11.2	3.25	0.019
RGB (ResNet-18)	RGB	11.21	42.8	28.6	5.67	0.032
RGB (ResNet-50)	RGB	23.50	89.7	80.2	4.56	0.026
RGB+DWT	RGB+DWT(L)	11.38	43.4	47.8	4.82	0.028
LAB (ResNet-18)	L*a*b*	11.21	42.8	36.7	4.56	0.026
Proposed LAB+DWT	L*a*b*+DWT(L)	11.38	43.4	44.7	5.40	0.031

Note: Training time, single fold, 30 epochs. Inference time: 175 test images, single-fold model. Model size = parameters × 4 bytes. Hardware and timing protocol as described above. Small differences between models may reflect implementation-level overhead.

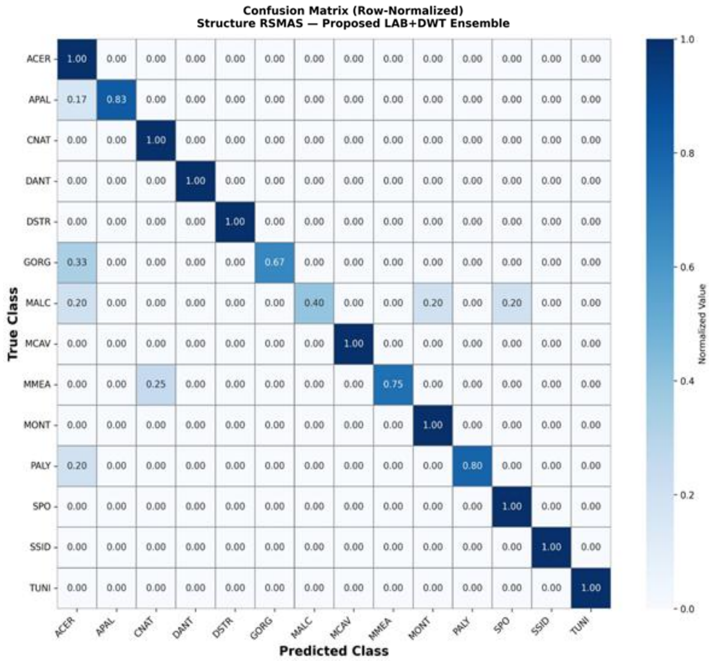
4.6. Confusion matrix analysis

Analysis of the row-normalized confusion matrices (Figure 3) reveals that classification reliability depends strongly on morphological distinctiveness. Species with unique structural or textural signatures exhibit high separability, while morphologically similar taxa remain prone to mutual confusion, reflecting known challenges in visual coral taxonomy. In Figure 3(a), the proposed LAB+DWT ensemble achieves perfect classification (100% recall) for 13 of 14 RSMAS species. The only misclassification involves *Acropora palmata* (APAL), where one sample is classified as *Acropora cervicornis* (ACER). Both species belong to the genus *Acropora* and share similar branching morphology, making this confusion biologically plausible. In Figure 3(b), the structure RSMAS results show that classification accuracy remains high overall (88.33%) but reveals interpretable error patterns concentrated among species with similar colony-level growth forms. *Millepora alcicornis* (MALC) exhibits the lowest recall, with misclassifications distributed across *Acropora cervicornis* (ACER) and Sponge (SPO), all of which can display similar massive or branching forms when viewed at colony scale. *Gorgonians* (GORG) show confusion with ACER, attributable to shared branching structures. *Meandrina meandrites* (MMEA) is occasionally confused with *Colpophyllia natans* (CNAT), both of which exhibit encrusting or brain-like growth morphology.

Differences between the two datasets highlight the role of visual scale. In the RSMAS domain, errors are primarily driven by similarities in fine-scale texture and pigmentation patterns, whereas in the structure RSMAS domain, confusion shifts toward taxa with similar colony-level growth forms. Nevertheless, certain class pairs—particularly MALC–ACER and CNAT–MMEA—remain difficult to separate across all model configurations, indicating that these classification ambiguities arise from intrinsic biological similarity rather than purely algorithmic limitations.



(a)



(b)

Figure 3. Row-normalized confusion matrices of the proposed LAB+DWT ensemble: (a) RSMAS test set and (b) Structure RSMAS test set. Diagonal values indicate per-class recall. Off-diagonal entries indicate the proportion of misclassified samples relative to the true class

4.7. Limitations

Several limitations of this study should be noted. First, the evaluation is based on two related datasets with a limited total sample size, and the structure RSMAS test set contains only 60 images. Therefore, domain-level accuracy values should be interpreted with awareness of the discrete 1.67 percentage-point step size per sample. With such a small test set, a single additional misclassification changes accuracy by 1.67 percentage points, which limits the precision of domain-gap comparisons. Second, no independent reef-site validation was included; both datasets originate from Caribbean reef environments, and generalization to Indo-Pacific or other biogeographic regions remains untested. Third, the DWT parameters (db2 wavelet, single-level decomposition, $\alpha = 0.5$) were selected based on preliminary experiments rather than exhaustive optimization. Early trials using full-image FFT-based frequency representations did not produce meaningful classification performance, which motivated the use of spatially localized wavelet decomposition. However, alternative wavelet families or multi-level decompositions were not systematically compared and may yield different results. Fourth, the current evaluation does not include comparison with vision transformers (ViT) or attention-based fusion mechanisms, which have shown strong performance in other fine-grained recognition tasks. Finally, all inference timing was measured using a CPU-only MATLAB implementation; GPU-accelerated deployment would yield different runtime characteristics. Future research may also investigate transformer-based architectures for coral classification. ViT models have demonstrated strong capability in modeling long-range spatial relationships and may provide additional benefits for colony-level imagery, where global structural information is particularly important [32].

5. CONCLUSION

This study investigated coral image classification under two imaging regimes: texture-dominant close-range imagery and morphology-focused colony-level imagery. RGB-based networks achieve high accuracy in texture-rich conditions but exhibit reduced performance when discriminative cues shift toward colony-level morphology, indicating a dependence on fine-scale texture features. To address this limitation, a dual-branch framework integrating CIELAB color representation with DWT-derived structural descriptors was proposed. By separating luminance and chromatic components and incorporating high-frequency structural information through Daubechies-2 wavelet decomposition, the architecture introduces complementary feature representations prior to convolutional abstraction.

The proposed model requires only 1.6% additional parameters over the single-branch ResNet-18 baseline (11.38M vs. 11.21M). Under the single best-fold evaluation protocol, the proposed LAB+DWT model achieved 96.52% accuracy on RSMAS and 88.33% accuracy on structure RSMAS, reducing the domain gap from 22.46% for the RGB ResNet-50 baseline to 8.19%, indicating improved generalization while maintaining a compact computational profile.

Analysis of confusion matrices shows that remaining misclassifications are primarily associated with morphologically similar taxa. Species with distinctive structural characteristics exhibit high separability, while massive and encrusting corals with comparable growth forms remain more difficult to distinguish, suggesting that certain ambiguities arise from intrinsic biological similarity rather than purely algorithmic limitations.

Although the datasets used in this study cover multiple coral taxa and imaging conditions, further evaluation on additional reef environments, geographic regions, and larger-scale datasets would strengthen the generalizability of the proposed framework. Future work may also investigate more advanced feature extraction and fusion strategies, including multi-level wavelet decomposition, learnable wavelet filters, and adaptive feature fusion mechanisms, to further improve robustness for morphologically ambiguous coral species.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Faculty of Science, Chulalongkorn University, for institutional support and facilities that made this research possible. The authors also thank colleagues and staff members for their assistance and constructive feedback during the preparation of this work.

FUNDING INFORMATION

This research received no external funding.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Pracharat Sa-ngadsup	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓		✓	
Chawan Koopipat	✓	✓	✓			✓	✓	✓	✓	✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author (P. Sa-ngadsup) upon reasonable request.

REFERENCES

[1] D. Souter, S. Planes, J. Wicquart, M. Logan, D. O. Obura, and F. Staub, “Status of coral reefs of the world: 2020 chapter 2. status of coral reefs of the world,” GCRMN. [Online]. Available: https://www.unep.org/cobsea/resources/reports/status-coral-reefs-world-2020

[2] T. P. Hughes *et al.*, “Global warming and recurrent mass bleaching of corals,” *Nature*, vol. 543, no. 7645, pp. 373–377, Mar. 2017, doi: 10.1038/nature21707.

[3] M. González-Rivero *et al.*, “Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis,” *Remote Sensing*, vol. 8, no. 1, p. 30, Jan. 2016, doi: 10.3390/rs8010030.

[4] A. Apprill *et al.*, “Toward a new era of coral reef monitoring,” *Environmental Science and Technology*, vol. 57, no. 13, pp. 5117–5124, Apr. 2023, doi: 10.1021/acs.est.2c05369.

[5] P. L. Jokiel *et al.*, “Comparison of methods used to estimate coral cover in the Hawaiian Islands,” *PeerJ*, vol. 3, no. MOA 2005-008/6882, p. 22 pp, 2005, [Online]. Available: http://scholar.google.com/scholar?start=0&hl=en&as_sdt=0,5&cluster=1284270061445928239.

[6] M. S. A. C. Marcos, M. N. Soriano, and C. A. Saloma, “Classification of coral reef images from underwater video using neural networks,” *Optics Express*, vol. 13, no. 22, p. 8766, 2005, doi: 10.1364/opex.13.008766.

[7] A. Ganesan and S. M. Santhanam, “A novel feature descriptor based coral image classification using extreme learning machine with ameliorated chimp optimization algorithm,” *Ecological Informatics*, vol. 68, p. 101527, May 2022, doi: 10.1016/j.ecoinf.2021.101527.

[8] A. Lumini, L. Nanni, and G. Maguolo, “Deep learning for plankton and coral classification,” *Applied Computing and Informatics*, vol. 19, no. 3–4, pp. 265–283, Jun. 2023, doi: 10.1016/j.aci.2019.11.004.

[9] A. S. M. Shihavuddin, N. Gracias, R. Garcia, A. C. R. Gleason, and B. Gintert, “Image-based coral reef classification and thematic mapping,” *Remote Sensing*, vol. 5, no. 4, pp. 1809–1841, Apr. 2013, doi: 10.3390/rs5041809.

[10] A. P. R. Nababan, T. Haryanto, and S. H. Wijaya, “Classification of coral images using support vector machine with gray level co-occurrence matrix feature extraction,” *International Journal on Informatics Visualization*, vol. 9, no. 3, pp. 936–946, May 2025, doi: 10.62527/joiv.9.3.2708.

[11] T. N. T. Arsad, E. A. Awalludin, Z. Bachok, W. N. J. H. W. Yussof, and M. S. Hitam, “A review of coral reef classification study using deep learning approach,” in *AIP Conference Proceedings*, 2023, p. 050005, doi: 10.1063/5.0110245.

[12] A. Mahmood *et al.*, “Deep Learning for Coral Classification,” in *Handbook of Neural Computation*, Elsevier, 2017, pp. 383–401. doi: 10.1016/B978-0-12-811318-9.00021-1.

[13] B. Reshma, B. Rahul, K. R. Sreenath, K. K. Joshi, and G. Grinson, “Taxonomic resolution of coral image classification with convolutional neural network,” *Aquatic Ecology*, vol. 57, no. 4, pp. 845–861, Dec. 2023, doi: 10.1007/s10452-022-09988-0.

[14] A. Alshahrani, H. Ali, E. Saif, M. Alsayed, and F. Alshareef, “Classification of coral reef species using computer vision and deep learning techniques,” *Engineering, Technology and Applied Science Research*, vol. 14, no. 5, pp. 16478–16485, Oct. 2024, doi: 10.48084/etasr.8044.

[15] A. Gómez-Ríos, S. Tabik, J. Luengo, A. S. M. Shihavuddin, and F. Herrera, “Coral species identification with texture or structure images using a two-level classifier based on convolutional neural networks,” *Knowledge-Based Systems*, vol. 184, 2019, doi: 10.1016/j.knosys.2019.104891.

[16] X. Shao *et al.*, “Deep learning for multilabel classification of coral reef conditions in the indo-pacific using underwater photo transect method,” *Aquatic Conservation: Marine and Freshwater Ecosystems*, vol. 34, no. 9, Sep. 2024, doi: 10.1002/aqc.4241.

[17] H. Nurill-Nabilla and W. N. Mohd-Isa, “Deep or shallow networks: coral types classification using residual network models,” in *Lecture Notes in Electrical Engineering*, vol. 1183 LNEE, 2024, pp. 77–89. doi: 10.1007/978-981-97-2007-1_7.

[18] S. Anwar and C. Li, “Diving deeper into underwater image enhancement: a survey,” *Signal Processing: Image Communication*, vol. 89, p. 115978, Nov. 2020, doi: 10.1016/j.image.2020.115978.

[19] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, “WaterGAN: unsupervised generative network to enable real-time color correction of monocular underwater images,” *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 387–394, 2018, doi: 10.1109/LRA.2017.2730363.

[20] C. Y. Li, J. C. Guo, R. M. Cong, Y. W. Pang, and B. Wang, "Underwater image enhancement by Dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, Dec. 2016, doi: 10.1109/TIP.2016.2612882.
[21] X. Huang, Y. Sun, X. Zhang, and J. Zhang, "Underwater image enhancement by combining LAB color space and double-opponency mechanism," *Journal of Internet Technology*, vol. 26, no. 6, pp. 775–784, Nov. 2025, doi: 10.70003/160792642025112606006.
[22] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, Jul. 1989, doi: 10.1109/34.192463.
[23] J. Amin, M. Sharif, N. Gul, M. Yasmin, and S. A. Shad, "Brain tumor classification based on DWT fusion of MRI sequences using convolutional neural network," *Pattern Recognition Letters*, vol. 129, pp. 115–122, Jan. 2020, doi: 10.1016/j.patrec.2019.11.016.
[24] J. M. Fortuna-Cervantes *et al.*, "Texture and materials image classification based on wavelet pooling layer in CNN," *Applied Sciences (Switzerland)*, vol. 12, no. 7, p. 3592, Apr. 2022, doi: 10.3390/app12073592.
[25] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," *Advances in Neural Information Processing Systems*, vol. 2017-December, pp. 4468–4476, 2017, doi: 10.5555/3294996.3295200.
[26] M. Sun, Y. Yuan, F. Zhou, and E. Ding, "Multi-attention multi-class constraint for fine-grained image recognition," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11220 LNCS, pp. 834–850, 2018, doi: 10.1007/978-3-030-01270-0_49.
[27] X. Zhao, P. Huang, and X. Shu, "Wavelet-attention CNN for image classification," *Multimedia Systems*, vol. 28, no. 3, pp. 915–924, Jun. 2022, doi: 10.1007/s00530-022-00889-8.
[28] T. Williams and R. Li, "Advanced image classification using wavelets and convolutional neural networks," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Dec. 2017, pp. 233–239, doi: 10.1109/icmla.2016.0046.
[29] J. Li, L. Zhao, H. Li, X. Xue, and H. Liu, "MixRformer: dual-branch network for underwater image enhancement in wavelet domain," *Sensors*, vol. 25, no. 11, p. 3302, May 2025, doi: 10.3390/s25113302.
[30] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, Oct. 1988, doi: 10.1002/cpa.3160410705.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[32] A. Dosovitskiy *et al.*, "An image is worth 16X16 words: transformers for image recognition at scale," in *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.

BIOGRAPHIES OF AUTHORS

Pracharat Sa-ngadsup is a lecturer with the Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand. He received a Ph.D. degree in Imaging Science and Printing Technology. Prior to joining academia, he worked in the camera manufacturing industry, gaining experience in imaging systems and camera technology. His research interests include image processing, deep learning, pattern recognition, and color management. His recent research focuses on marine ecosystem monitoring, particularly coral reef classification using RGB-IR imaging and shape-based features. His work also involves underwater photogrammetry and remote sensing techniques for coral reef mapping, structural monitoring, and environmental assessment. His broader research interests include applied computer vision and imaging technologies for environmental analysis. He can be contacted at email: pracharat.sa@chula.ac.th.

Chawan Koopipat received the Ph.D. degree in Image Information from Chiba University, Japan. He is currently an Associate Professor with the Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand. His research interests include imaging science, color imaging, printing technology, and computational image analysis. His work also involves imaging applications for environmental monitoring, including coral reef observation using multispectral and RGB-IR imaging techniques. He has contributed to research in imaging technology and computer vision for scientific and industrial applications. He can be contacted at email: chawan.k@chula.ac.th.