

A Comparative Evaluation of Hybrid Fusion Strategies for Turtle Identification

Penilaian Perbandingan Strategi Gabungan Hibrid untuk Pengecaman Penyu

*Mohd Erman Safawie Che Ibrahim¹, Wan Nural Jawahir Hj Wan Yussof*¹, Muhammad Suzuri Hitam¹, Ezmahamrul Afreen Awalludin², Mohamad Fathullah Ruslan³, Siti NurFarahim Shahaudin¹*

¹*Group of Artificial Intelligent, Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

²*Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia*

³*Fisheries Research Institute Rantau Abang, 23050 Kuala Dungun, Terengganu, Malaysia*

**Corresponding author: wannurwy@umt.edu.my*

Received 12 October 2025

Accepted 18 April 2026, Available online 30 June 2026

ABSTRACT

This paper presents a comparative study of Coiflet1 wavelet- and Gabor-based fusion strategies combined with deep learning for individual sea turtle identification. The visual identification of marine turtles continues to be challenging because of the small amount of annotated data, high inter class similarity and class imbalance. To achieve this, a feature to image transformation framework is presented in which statistical descriptors obtained from Coiflet1 wavelet transforms and Gabor filters are restructured into surrogate image representations and then fed into a pre-trained ResNet-18 convolutional neural network. Four fusion schemes are applied which are Coiflet1 only, Gabor only, sequential Coiflet1 and Gabor, and parallel Coiflet1 and Gabor. All models are tested on a custom dataset that consists of 1,426 labeled images of turtles in 20 different identities. We evaluate the performance using several evaluation metrics, including accuracy, precision, recall, F1-score, ROC-AUC, PR-AUC, Cohen kappa, and training time. Experimental results have shown that all the models show impressive performance (accuracies more than 93%). The Coiflet1 only configuration outperforms the others with 97.57% accuracy and a Cohen's kappa of 0.97. It also achieves the best precision (99.12%), recall (94.83%), and F1 score (96.37%), while the Gabor-only (96.53% accuracy) and hybrid fusion configurations both trails. These results show that if structured as surrogate images that independently reformulate handcrafted wavelet features, it does better than those in more complex fusion schemes in resource-constrained deep learning. The proposed framework sets a new quantitative level for hybrid wavelet-CNN architectures and provides insight into the trade-offs among fusion complexity, model generalization and computational efficiency.

Keywords: Deep Learning; Wavelet Transform; Gabor Filters; Feature Fusion; Ecological Monitoring

ABSTRAK

Kertas ini membentangkan kajian perbandingan strategi gabungan berasaskan wavelet Coiflet1 dan Gabor yang digabungkan dengan pembelajaran mendalam untuk pengenalanpastian penyus laut individu. Pengenalanpastian visual penyus marin terus mencabar kerana jumlah data beranotasi yang kecil, persamaan antara kelas yang tinggi dan ketidakseimbangan kelas. Untuk mencapai matlamat ini, rangka kerja transformasi imej ciri dibentangkan di mana deskriptor statistik yang diperoleh daripada transformasi wavelet Coiflet1 dan penapis Gabor disusun semula menjadi perwakilan imej pengganti dan kemudian dimasukkan ke dalam rangkaian saraf konvolusi ResNet-18 yang telah dilatih terlebih dahulu. Empat skema gabungan digunakan iaitu Coiflet1 sahaja, Gabor sahaja, Coiflet1 dan Gabor berjajukan, dan Coiflet1 dan Gabor selari. Semua model diuji pada set data tersuai yang terdiri daripada 1,426 imej berlabel penyus dalam 20 identiti berbeza. Kami menilai prestasi menggunakan beberapa metrik penilaian, termasuk ketepatan, kejituan, ingatan, skor F1, ROC-AUC, PR-AUC, Cohen kappa, dan masa latihan. Keputusan eksperimen menunjukkan bahawa semua model menunjukkan prestasi yang mengagumkan (ketepatan lebih daripada 93%). Konfigurasi Coiflet1 sahaja mengatasi yang lain dengan ketepatan 97.57% dan Cohen's kappa sebanyak 0.97. Ia juga mencapai ketepatan terbaik (99.12%), ingatan semula (94.83%), dan skor F1 (96.37%), manakala konfigurasi Gabor sahaja (ketepatan 96.53%) dan gabungan hibrid kedua-duanya ketinggalan. Keputusan ini menunjukkan bahawa jika distrukturkan sebagai imej pengganti yang merumus semula ciri wavelet buatan tangan secara bebas, ia berfungsi lebih baik daripada yang terdapat dalam skema gabungan yang lebih kompleks dalam pembelajaran mendalam yang terhad sumber. Rangka kerja yang dicadangkan menetapkan tahap kuantitatif baharu untuk seni bina wavelet hibrid-CNN dan memberikan pandangan tentang pertukaran antara kerumitan gabungan, generalisasi model dan kecekapan pengiraan.

Kata kunci: Pembelajaran Mendalam; Transformasi Gelombang; Penapis Gabor; Penggabungan Ciri; Pengecaman Hidupan Liar

INTRODUCTION

Marine turtles are one class of the most important and endangered species that are essential to maintaining marine ecological health and stability (Hays et al., 2025). Species level individual turtle identification is recognized as crucial in conservation biology as it would facilitate monitoring in the long term, population estimation, tracking of migrations, and study of behavior (Panarese et al., 2025). Typical identification methods are flipper tagging and passive integrated transponder (PIT) and photo identification methods. So far, they are helpful techniques but suffer from physical deterioration, tag loss, human error, as well as practical limits when they are utilized in the real world (Neves-Ferreira et al., 2023). Computer vision and deep learning applications have recently been used to automate turtle visual identification systems, enabling improved reliability and scalability of identification (Adam et al., 2025; Jeantet et al., 2024). While it makes large-scale CNNs extremely robust to achieve different species of identification in wildlife, their successes are limited to underwater turtle identification where they suffer from very low visibility and color distortion, varied illumination, and morphological similarity among individuals (Stavelin et al., 2016; Ding et al., 2024). Additionally, due to data scarceness and class imbalance in datasets, it is further a struggle for model generalization in small scale datasets focusing on a single species (Alzubaidi et al., 2023). Common deep learning methods often depend on large and sufficiently balanced datasets to discriminate features (LeCun et al., 2015). On the other hand, handcrafted descriptors have the far-higher representation power in low data regimes, especially for texture

rich and fine-grained identification tasks (Alshazly et al., 2019; Bento et al., 2022). These represent multiresolution and orientation specific information that can supplement representations of CNN-based approaches.

Hybrid structures integrating handcrafted statistical descriptors with deep feature representations have been gaining attention recently as they enhance identification performance and robustness (Ullah et al., 2025; Tschuchnig et al., 2025). Of these handcrafted approaches, Coiflet1 wavelet transforms and Gabor filters have made remarkable contributions towards the multiscale texture and frequency domain information detection. Coiflets are discrete wavelets designed by Ingrid Daubechies with near symmetric properties (Daubechies, 1992; Santos et al., 2014; Bhati et al., 2017). The Coiflet1 wavelet family is famous for its time frequency localization in sequence with symmetry properties that allow for feature extraction on a rich multiscale scale. Gabor filters with outstanding orientation and spatial frequency selective ability were adopted through biological inspiration to model the human visual system response to edge and pattern information (Daugman, 1985; Lee, 1996). They are both handcrafted descriptors that can be exploited in CNN architecture at feature level and/or decision level to further improve CNN model interpretability and accuracy under resource limited and data sparse contexts (Tang et al., 2022; Zhao et al., 2024).

Motivated by these advantages, this study presents a comparative evaluation of Coiflet1 wavelet-based and Gabor-based fusion strategies integrated with deep learning for individual sea turtle identification. Specifically, the study investigates the effectiveness of wavelet-only, Gabor-only, and hybrid fusion strategies in both sequential and parallel configurations to enhance recognition accuracy and model generalization. The proposed fusion techniques are then methodically confirmed with a dedicated dataset of individually labelled turtle images, and we set a quantitative reference point for the development of hybrid wavelet-CNN architectures in fine-scale marine animal recognition. It also presents the trade-offs between fusion complexity, model generalization, and computational cost and thus promotes knowledge for future development of efficient deep hybrid systems for underwater visual recognition.

RELATED WORK

Accurate individual identification is an important issue in marine conservation and ecological monitoring as it determines population sizes, behaviour and longterm status of species. Methods for identifying turtles such as tagging, matching scute patterns and pictorial identification with photographs were primarily used to monitor turtle populations (Tabuki et al., 2021; Wallace et al., 2010). However, existing approaches are currently limited by tag loss, image discrepancies and human error hampering the reliability of the long-term datasets and field observations (Neves-Ferreira et al., 2023). Recently, it has evolved into automated image-based identification systems driven by computer vision and artificial intelligence whereby marine species can be tracked efficiently and effectively without invasion in a scalable fashion (Hays et al., 2018; Panarese et al., 2025).

Deep learning methods, particularly convolutional neural networks (CNNs), are used in individual turtle identification and other wildlife recognition tasks (Jeantet et al., 2024; Adam et al., 2025). Although these methods effectively identify images, underwater visual identification is still challenging because of poor visibility, illumination variation, color distortion and background interference, all of which severely affect the image quality and model performance (Schettini et al., 2010; Hu et al., 2022). Also, marine image datasets have a limited size and severe class imbalance, which restrict the generalization capability of deep neural

networks on fine-grained classification tasks (De Zarzà et al., 2023). Traditional CNN architectures usually require large and balanced datasets to learn discriminative features effectively (LeCun et al., 2015; Simonyan and Zisserman, 2014) and these limitations push for the study of complementary feature extraction approaches.

Prior to the increased use of deep learning, handcrafted feature descriptors like Scale Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), Gabor filters, and wavelet transforms had been broadly used in pattern recognition and texture analysis (Daugman, 1985; Lee, 1996). As such, these descriptors encode multi-scale and orientation sensitive information which are more appropriate for texture rich biological surfaces as zebra stripes and cheetah spots, among many more, for individual recognition (Lahiri et al., 2024; Vidal et al., 2021). Of these methods, wavelet transforms have garnered much focus in the field of modelling due to their ability to capture localized frequency information and multi resolution texture representations (Mallat, 2002). Most importantly, the Coiflet family of orthogonal wavelets has near symmetry and compact support which provide better time frequency localization and energy compaction ability than other wavelet families like in standard design such as Haar and Daubechies (Daubechies, 1992). Akbar (2020) proved the capabilities of Coiflet wavelets for image compression and noted that Coif1 had the lowest compression ratio and Coif5 had the highest reconstruction fidelity. With these properties, Coiflet1 is especially important for fine-grained detection of structural variations in a very subtle way under different environmental conditions that are needed for specific recognition tasks.

In much the same way, Gabor filters have been widely used for texture study because of their strong orientation and spatial frequency selectivities based on the response properties of the human visual cortex (Daugman, 1985; Lee, 1996). Having high stability against rotation, scale, and illumination differences, they are effective for face recognition, iris recognition, and underwater object identification (Kamarainen et al., 2006; Riaz et al., 2018). Gabor descriptors have shown strong local texture encoding in underwater imaging situations requiring subtle pattern differences for differentiating individuals, and these can be a good supplement to wavelet-based representations (Abu and Diamant, 2023).

In order to overcome the limitations of individual deep learning and handcrafted methods, hybrid feature fusion approaches have seen their way to the forefront of the literature and are considered an effective alternative by combining handcrafted descriptors with deep neural representations (Artabaz and Sliman, 2025; Tschuchnig et al., 2024). Thus, feature-level fusion can be undertaken to integrate handcrafted descriptors into CNN extracted embeddings featuring low-level texture types with high-level semantic features simultaneously (Can and Aydin, 2025). Mahmood et al. (2016), for example, combined handcrafted texture and color descriptors with multi-scale CNN representations to realize the state-of-the-art coral reef classification performance. Similarly, Fujieda et al. (2017) proposed wavelet convolutional neural networks for texture classification by incorporating spectral wavelet analysis into CNN architectures, demonstrating that frequency domain information can complement spatial domain CNN processing. In marine classification tasks, Hu et al. (2012) developed wavelet-based texture feature extraction models from fish images, achieving 97.96% classification accuracy but with some classes misclassified due to texture similarity between species.

Apart from direct feature fusion, it has been investigated in recent years to transform non-image statistical data into structured image representations for CNN processing. Sharma et al. (2019) proposed a way to extract and transform non-image samples into organized image forms for input to CNN. Zhu et al. (2021) used a systematic approach to the same problem by reorganizing

tabular features into pixel-based image layouts. The result preserves feature proximity relationships. More recently, Alenizy and Berri (2025) suggest making synthetic 2D images of tabular data through mathematical transformations such as rotation, translation, and reflection to simulate the spatial relationships among features. These methods enable CNN architectures like ResNet and EfficientNet to handle statistical descriptors in a unified visual domain and exploit hierarchical feature extraction mechanisms.

Whilst it has also been shown that wavelet transform, Gabor filters and hybrid CNN frameworks can effectively work independently in previous literature, comparative research methods for fusion based on handcrafted descriptors and deep learning approaches are still limited. Current research tends to be limited to either isolated handcrafted descriptors or direct CNN-based classification on the raw image data, neglecting the structured texture level fusion strategies. Moreover, empirical analyses of a number of fusion schemes including sequential as well as parallel combinations, have not been extensively studied under fine-grained identification scenarios. To bridge these gaps, this work makes three key contributions:

1. Comparative evaluation of hybrid fusion strategies based on Coiflet1 wavelet, Gabor filters and CNNs.
2. Feature to image transformation pipeline which allows efficient integration of handcrafted descriptors with deep neural networks.
3. Extensive benchmarking on a custom turtle image dataset to demonstrate robustness and effectiveness under fine-grained, low data, and resource constrained conditions.

METHODOLOGY

This study uses four fusion strategies, namely by combining handcrafted features based on Coiflet1 and Gabor wavelets with a deep learning architecture for individual sea turtle identification. The overall workflow, a detailed overview detailing the systematic pipeline adopted from image acquisition to model evaluation focusing on repeatability, interpretability and consistency with ResNet-18, is illustrated by Figure 1. The pipeline is organized around six stages that include image data store preparation, preprocessing, feature extraction and transformation, fusion strategy design, deep learning integration using ResNet-18 and performance evaluation. Each stage is carefully structured to take into account the contribution that handcrafted and hybrid feature representations can make to classification performance in data-limited scenarios.

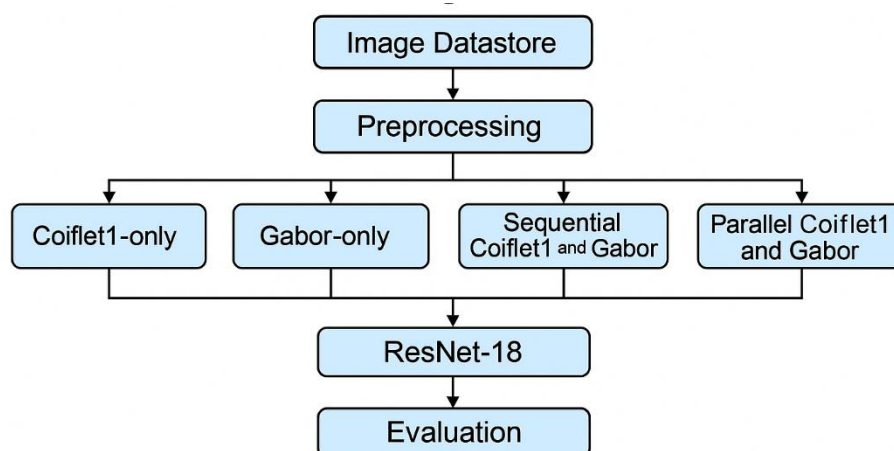


FIGURE 1. Flow diagram of the turtle identification pipeline with fusion strategy.

We chose ResNet-18 as the deep learning model for this work, to ensure a balanced evaluation and assessment of all fusion methods. Furthermore, ResNet-18 also has fewer trainable parameters than architectures such as ResNet-50 and DenseNet and thus reduces the risk of overfitting when trained on small datasets, while it can guarantee competitive classification performance through the inclusion of residual learning (He et al., 2016). The lightweight nature of this architecture allows for faster training and inference, which allows for the evaluation of various fusion arrangements in a controlled setting. ResNet-18 can also be considered a useful standard backbone because this study primarily compares the performance of hand-crafted feature fusion strategies rather than deep refinement on the network.

DATA PREPROCESSING

The dataset used in this study consists of 1,426 images from 20 turtle identities, designated T1 up to T20. Nevertheless, class distributions are extremely unbalanced and the samples for each identity greatly differ. There are 286 examples in class T6 while only 11 in class T8. This imbalance represents an important difficulty for classification models as it requires the use of good feature representations together with a rigorous evaluation process to reduce the risk of biased learning to improve generalization. In order to ensure statistical integrity, the data was split into 80% training set and 20% test set with stratified random sampling to obtain proportional class representation for each turtle. Each image was standardized by stepwise processing. Specifically, we rescaled each image to 224×224 pixels to create consistent spatial quality and match the input conditions of the CNN architectures. The class distribution of the 20 turtle identities is shown in Table 1, demonstrating that the sample allocation is skewed, which potentially creates bias in training.

TABLE 1. Class distribution of turtle dataset across 20 individual identities.

Class ID	Sample Count	Class ID	Sample Count
T1	219	T11	28
T2	138	T12	16
T3	13	T13	13
T4	121	T14	28
T5	114	T15	96
T6	286	T16	14
T7	63	T17	19
T8	11	T18	24
T9	14	T19	21
T10	25	T20	163

The training set was augmented for better learning of future generations to improve modeling generalization and prevent overfitting. The augmentation strategy consists of horizontal reflections and random rotations within ± 10 degrees, random isotropic scaling between 0.9 and 1.0. These are adjustments that alter the appearances in a subtle way, but are cognizable at all times. The obtained data from each turtle's characteristics becomes more potent against variation, posed and scaled. In Figure 2, the great natural variability in pose, illumination, and underwater imaging conditions within which the dataset was based is presented, which provides a reasonable benchmark in estimating the robustness of the proposed methods for feature extraction at the same time. For texture cues related to individual discrimination, two-dimensional Coiflet-1 wavelet decompositions were generated, and Gabor magnitude filtering over the resulting wavelet subbands was carried out by using a multi-orientation, multi-scale

filter bank. These analyses provide fine-grained texture profiles (possibly spatial frequency distributions that account for the discriminative separability among individual turtles) that can be derived.

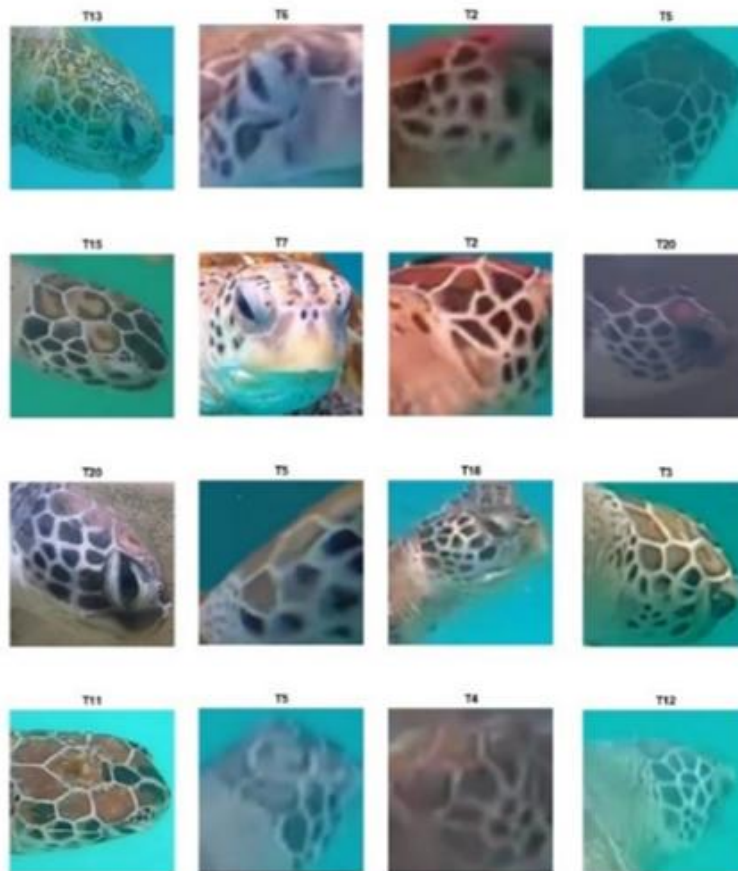


FIGURE 2. Sample images from the turtle dataset across different individuals.

FEATURE EXTRACTION AND SURROGATE TRANSFORMATION

The concept of key feature extraction and surrogate transformation provides a pathway for traditional signal processing techniques and deep learning-based feature learning within the proposed framework. Rather than serving the raw images as input to a CNN directly, the hand-crafted statistical descriptors are transformed into surrogate grayscale images, which allow the network to interpret structured statistical patterns in spatially coherent form. It is based on the interpretability of wavelet and Gabor features and the discriminative capacity of deep architectures (Lahmiri and Boukadoum., 2013; Devulapalli et al., 2023; Nakka et al., 2023). Formally, given a grayscale image $I(x, y) \in \mathbb{R}^{M \times N}$, the objective is to transform I into a surrogate image representation $S(x, y)$ such that:

$$S(x, y) = \mathcal{T}(\Phi(I(x, y)))$$

Where $\Phi(\cdot)$ is the handcrafted feature extraction operator, and $\mathcal{T}(\cdot)$ denotes the normalization and tiling process which maps extracted statistical descriptors into an image domain of fixed dimensions 224×224 . This Coiflet1 wavelet transform gives a highly localized multi resolution representation of the input image by decomposing it into sub-bands containing distinct spatial frequency components. This decomposition was done via the discrete wavelet transform

(DWT), defined as (Mallat, 1989)

$$W_{\psi}(a, b) = \sum_x \sum_y I(x, y) \psi_{a,b}(x, y)$$

Where $\psi_{a,b}(x, y) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b_x}{a}, \frac{y-b_y}{a}\right)$ is the scaled and translated version of the mother wavelet $\psi(x, y)$, and a, b represent scale and translation parameters, respectively. For the single level Coiflet1 decomposition, the image is divided into four sub-bands. Figure 3 displays the single level discrete wavelet transform (DWT) decomposition applied to a turtle facial image using the coiflet1 wavelet, representing approximation, horizontal, vertical, and diagonal details, respectively. From each sub-band, each descriptor vector is then normalized and tiled into a structured 4×5 grid, which is interpolated and resized to a 224×224 grayscale image, producing the Coiflet1 surrogate image. This transformation enables subsequent CNN layers to interpret handcrafted statistical features as spatially structured inputs while preserving hierarchical relationships across scales. Figure 4 presents Gabor magnitude responses obtained by convolving the wavelet subbands with a bank of multi orientation and multi scale Gabor filters.

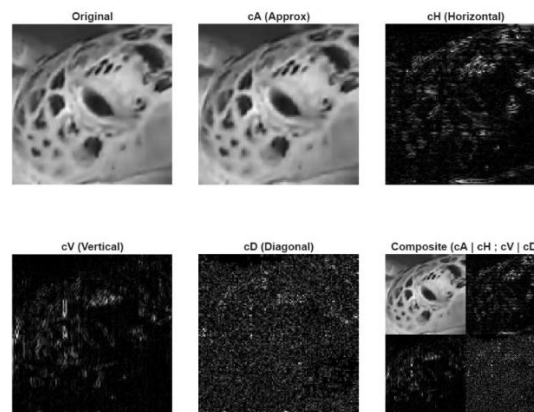


FIGURE 3. Single Level Coiflet1 Wavelet Decomposition of Turtle Facial Scute Texture.

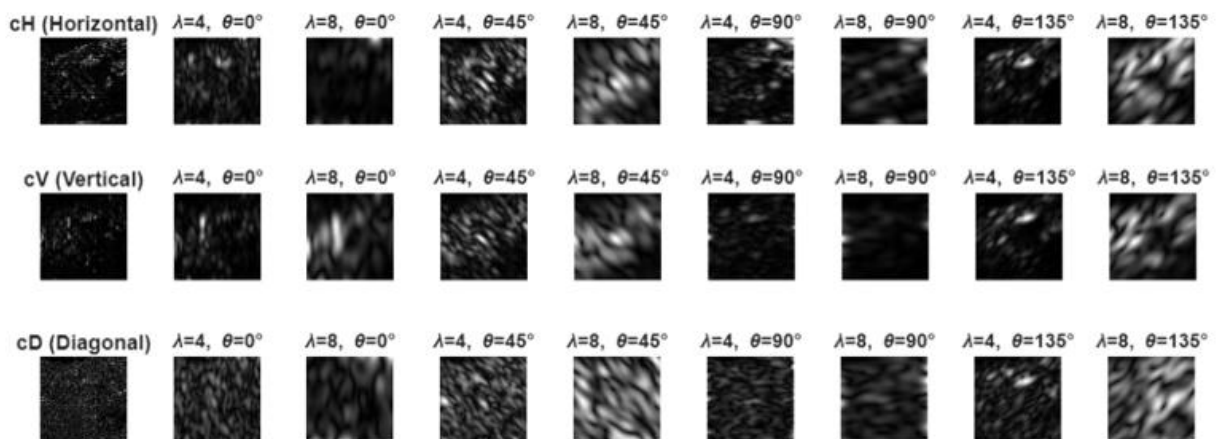


FIGURE 4. Gabor Magnitude Responses Across Coiflet1 Wavelet Subbands

The Gabor filter bank models the receptive field of simple cells in the human visual cortex,

providing orientation and spatial frequency selectivity (Daugman, 1985; Lee, 1996). A 2D Gabor filter is defined as:

$$G(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)$$

Where:

$$\begin{cases} x' = x \cos \theta + y \sin \theta \\ y' = -x \sin \theta + y \cos \theta \end{cases}$$

The period of the sinusoidal carrier is specified by the wavelength λ , so λ determines the filter's preferred spatial frequency; smaller λ increases sensitivity to fine textures; larger values emphasize coarser structures. The orientation θ gives kernel rotation with respect to the axes of the image so that when a certain angle is desired, the filter's response should be aligned to edges or ridges. The phase offset ψ shifts the sinusoidal component, determining whether the filter will be even or odd. The standard deviation σ of the Gaussian envelope determines the extent of the kernel's spacing, thereby achieving a balance between localization in the image plane and frequency selectivity. Last, the spatial aspect ratio γ determines the ellipticity of the Gaussian window, compressing the kernel on one axis in order to tune anisotropy: values $\gamma < 1$ produce elongated, orientation selective kernels whereas $\gamma \approx 1$ is close to circular symmetry. For reference, this study implemented a multi scale, multi orientation Gabor filter bank with scales of $\{2,4,8\}$ and the $0^\circ, 45^\circ, 90^\circ, 135^\circ$ orientations. Each image was convolved with the filter bank.

$$R_{s,\theta}(x, y) = I(x, y) * G(x, y; \lambda_s, \theta)$$

producing a set of magnitude response maps $\{|R_{s,\theta}(x, y)|\}$ that highlight texture and edge information across orientations. For each response map, the same five statistical descriptors (mean, standard deviation, skewness, kurtosis, and entropy) were extracted:

$$\Phi_{s,\theta} = [\mu_{s,\theta}, \sigma_{s,\theta}, \text{Skew}_{s,\theta}, \text{Kurt}_{s,\theta}, H_{s,\theta}]$$

The resulting descriptor matrix was normalized and mapped into a grayscale surrogate image $S_G(x, y)$ of dimension 224×224 :

$$S_G(x, y) = \mathcal{N}(\Phi_{s,\theta})$$

where $\mathcal{N}(\cdot)$ denotes min-max normalization followed by bicubic interpolation to match CNN input dimensions.

FUSION STRATEGIES

The fusion process aims to integrate handcrafted and learned features to improve identification performance under limited data conditions. The central premise of fusion in this research is that Coiflet1 wavelet and Gabor filter descriptors encapsulate complementary information in the spatial and frequency domains that can enrich the representational diversity of the deep learning feature space (Abdellatef et al., 2020; Artabaz and Sliman., 2025). Let $I(x, y)$ denote an input image of spatial dimensions $H \times W$. Two distinct handcrafted transformations are applied to $I(x, y)$ where Coiflet1 Wavelet Transform $W_c(I)$ capturing localized frequency components via multi resolution decomposition and gabor filter bank $G(I)$ to encoding orientation selective

texture responses through band pass filtering. Mathematically, the wavelet and Gabor feature maps can be represented as:

$$W_c(I) = \{f_{ij}^{(c)} | f_{ij}^{(c)} = \phi_c(I, s_i, d_i)\},$$

$$G(I) = \{g_{mn} | g_{mn} = \psi_c(I, \lambda_m, \theta_n)\}$$

where ϕ_c denotes the Coiflet1 wavelet basis at scale s_i and direction d_i , while ψ represents a Gabor kernel parameterized by wavelength λ_m and orientation θ_n . Each transform produces a set of coefficient matrices from which statistical descriptors such as mean μ , standard deviation σ , skewness γ , kurtosis κ , and entropy H are extracted as feature summaries:

$$F_W = [\mu_W, \sigma_W, \gamma_W, \kappa_W, H_W], \quad F_G = [\mu_G, \sigma_G, \gamma_G, \kappa_G, H_G].$$

These descriptors are normalized and converted into features to image surrogates, denoted S_W and S_G , each of dimension 224×224 . To systematically investigate the influence of fusion design on model generalization, four distinct fusion strategies were developed. Each strategy defines a function $\mathcal{F} : (S_W, S_G) \mapsto S_F$, where S_F is the fused surrogate image used as input to the pretrained CNN. Coiflet1-only replicated grayscale Coiflet1 surrogates across RGB channels.

$$S_F = \text{replicate}(S_W, 3)$$

where the single channel wavelet surrogate is replicated across the three RGB channels. This strategy emphasizes the multiresolution spatial frequency characteristics of the Coiflet1 transform alone. Gabor derived surrogates were replicated across all color channels, representing a texture-oriented baseline emphasizing orientation specific edge and pattern information.

$$S_F = \text{replicate}(S_G, 3)$$

like the coiflet but encoding purely orientation dependent textural information. Sequential Coiflet1 and Gabor Fusion surrogates were horizontally concatenated and resized to $224 \times 224 \times 3$. This spatial concatenation combines both modalities side by side enabling the CNN to learn from joint texture frequency regions.

$$S_F = \text{resize}([S_W || S_G], (224, 224, 3))$$

Where $[S_W || S_G]$ denotes the horizontal concatenation of Coiflet1 and Gabor surrogates, combining both descriptors sequentially along the width dimension before resizing. This approach attempts to unify global local feature patterns within a single spatial frame. Parallel coiflet1 and gabor fusion a channel level fusion was implemented by assigning Coiflet1 features to the red channel, Gabor features to the green channel, and their mean intensity map to the blue channel. This design allows simultaneous learning from both modalities within the same image while encoding cross modal correlations at the pixel level.

$$S_F = \text{cat}_3 \left(S_W, S_G, \frac{S_W + S_G}{2} \right)$$

in which Coiflet1 and Gabor surrogates are merged channel wise, assigning S_W to the red channel, S_G to the green channel, and its mean composite to the blue channel. Here the fusion also allows for increasing parallel frequency orientation integration, and the spatial

independence of the feature in each image can be retained, yet they can still be processed for inter channel correlation learning using CNN. These two fusion algorithms present a trade-off between feature complementarity and computational efficiency. Wavelet-based descriptors can be used to provide localized frequency information and are robust to illumination changes (Parmar et al., 2012), whereas Gabor filters have been found to improve the directional encoding of textures. In the next step these representations spatially unify to further increase intra-image diversity (sequential fusion), and parallel fusion takes advantage of RGB channels as complementary feature spaces to increase the CNN capacity to learn multi domain correlations. All the strategies were trained independently and based on a similar ResNet-18 backbone and hyperparameters, in order to conduct a fair comparison. Performance differences between the fusion modes are merely a result of current integration of representation, not any training heterogeneity (Yang et al., 2025).

RESULTS AND DISCUSSION

The feature analysis applied to image transformation strategies for turtle identification offers an interesting chance to investigate the trade-offs between handcrafted descriptors, fusion complexity, and deep learning.

COMPARATIVE STRATEGY PERFORMANCE

Analysis of the results of four handcrafted feature transformation strategies, Coiflet1 Wavelet Only, Gabor Only, Sequential Coiflet1 and Gabor Fusion, and Parallel Coiflet1 and Gabor Fusion, indicates variations in discriminative strength, model reliability, and computational efficiency. Table 2. Exhibited performance results of different feature and image fusion methods in ResNet-18. The Coiflet1 framework had the best performance overall with an accuracy of 97.57%, precision of 99.12%, and AUC of 99.99%, backed by a strong Cohens Kappa 0.97 showing excellent classifier agreement. This result shows that the wavelet based textural decomposition effectively captures structural and hierarchical information of turtle facial scute, which is stable across individuals. That it showed robustness in all metrics shows the sufficiency of Coiflet1 wavelets as standalone discriminative features, enabling ResNet-18 to extract separable deep representations even from handcrafted feature images. The Gabor only features also achieved high 96.53% performance, demonstrating their sensitivity to orientation and frequency although there was minor variability among the individuals. Accuracy decreased to 93.75% due to sequential fusion of Coiflet1 and Gabor, which indicates feature redundancy and information overlap that limits classifier generalization.

TABLE 2. Performance Comparison of Fusion Strategies Using ResNet-18

Fusion Strategy	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)	Cohen's Kappa	Training Time (s)
Coiflet1 Only	97.57	99.12	94.83	96.37	99.99	0.97	394.87
Gabor Only	96.53	97.52	95.28	96.12	99.98	0.96	394.00
Coiflet1 and Gabor (Sequential)	93.75	82.16	82.05	80.50	99.29	0.93	393.38
Coiflet1 and Gabor (Parallel)	96.18	95.42	91.56	92.26	99.95	0.96	375.81

In contrast, parallel fusion restored balanced performance 96.18% accuracy with improved

computational efficiency. Despite all methods achieving near perfect AUC more than 99%, these high ROC values may overstate performance under class imbalance, making F1-score and Cohen's Kappa more reliable for evaluation. Overall, the results confirm that handcrafted transformations particularly Coiflet1 and Gabor filters produce highly discriminative surrogate images for CNN-based ecological identification.

ROC CURVE ANALYSIS AND PERFORMANCE

ROC analysis of ResNet-18 on feature to image representations reveals nearly perfectly separable classes across the 20 turtle identities with AUC values consistently close to 1.0 in both Coiflet1 Wavelet and Gabor feature spaces. Confirmation of the strong discriminative capability and insignificant identity class overlap of the networks can be seen in the tightly clustered ROC curves along the top left axis. Importantly, Coiflet1-only and Gabor-only transformations produced uniformly maximal AUCs of approximately 1.00, revealing the capabilities of handcrafted descriptors to encode class distinct structural and spectral cues with minimal feature engineering. Figure 5 illustrates the ResNet-18 ROC curves obtained for the evaluated fusion strategies.

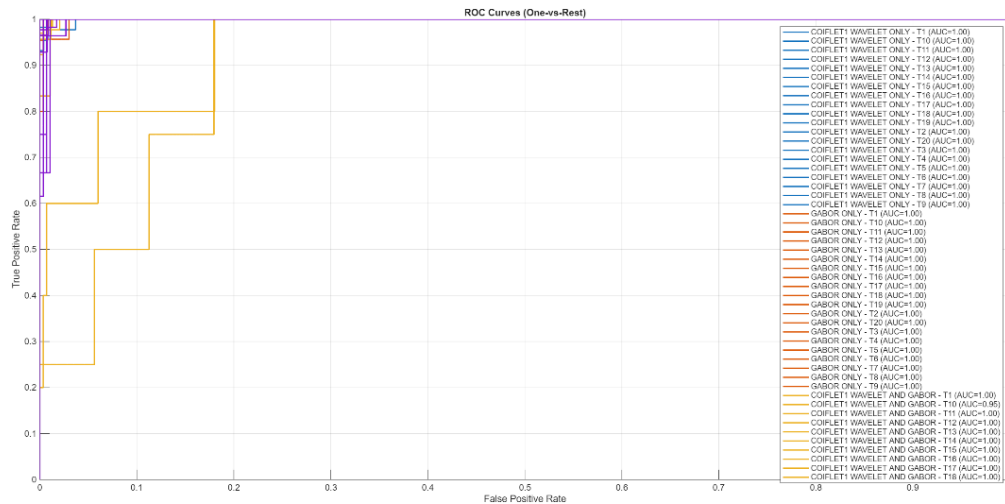


FIGURE 5. ROC Curves of ResNet-18

The slight AUC deviation (0.95) obtained for one sequential fusion case thus may suggest redundancy when combining heterogeneous filters and as such, could imply that fusion will be a matter of saturation rather than representational diversity increases. These findings altogether show that handcrafted feature transformations, if presented as surrogate images, are highly separable manifolds for CNN-based classification and support their inclusion at lightweight, interpretable front ends in deep ecological identification schemes. Overall, this comparative study presents the discriminative performance of the handcrafted representation under ResNet-18, in terms of accuracy, recall, and F1 score across feature to image approaches. Nevertheless, Coiflet1 Wavelet Only and Gabor Only configurations have a performance accuracy of over 96% that is near saturated indicating a capability of encoding stable spatial frequency textures specific to each turtle identity. While Sequential Coiflet-Gabor Fusion brought a marked drop in recall and F1-score to 82%, revealing that sequential layering introduces redundancy or conflicting features that reduce generalization. Parallel Fusion strategy was applied to alleviate part of the problem, presenting balanced and slightly reduced metrics indicating a moderate complementarity between spaces of features. We notice this as an overall consistency between accuracy and F1-score, leading to dependable confidence in the class model, while drop in fusion variants indicates

sensitivity to intra-class texture overlap. Comparative performance on accuracy, recall, F1-score is shown in Figure 6.

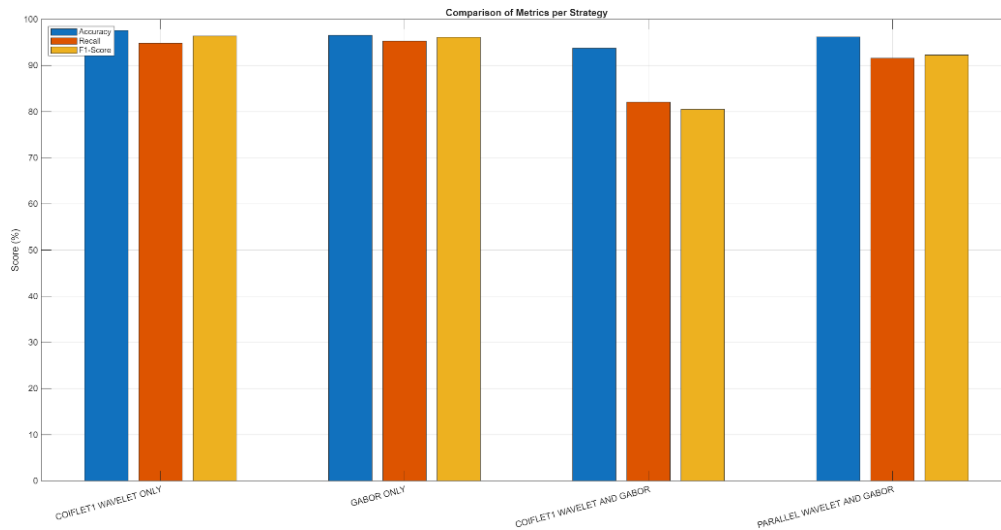


FIGURE 6. Comparative performance across accuracy, recall, and F1-score.

Taken together these results stress that single handcrafted descriptors are still most discriminatively informative, and fusion must be selectively designed to avoid representational redundancy in identification systems.

FEATURE CLUSTER VISUALIZATION

The PCA and t-SNE visualisations provide a comparison of feature separability with the four transformation strategies. For both 2D and 3D projections, the Coiflet1 Wavelet Only and Gabor Only configurations have well marked and compact clusters with less inter-class overlap, implying that feature encoding and discriminative quality are similar. In contrast, Sequential Coiflet and Gabor Fusion demonstrate a high degree of clustering and overlaps between classes, which reflect redundancy as a result of that excess and interaction of spatial frequency responses that leads to reduction of the feature orthogonality. Conversely, Parallel Fusion demonstrates moderate separability, meaning it preserves a bit of discriminative variance through concurrent representation learning. The t-SNE plots appear to confirm the finding, as we already see that single feature strategies have cleaner manifold boundaries than fusion-based strategies. Collectively, this data validates that handcrafted transforms, namely Coiflet1 wavelet and Gabor, create structured embeddings that are class-separating and can be effectively exploited in ResNet-18. Figure 7 Visualized PCA and t-SNE of fusion strategies. However, too much fusion lowers the compactness of latent features, indicating that best discriminability in representations of handcrafted to image results from keeping its functional features contiguous to avoid duplicate components.

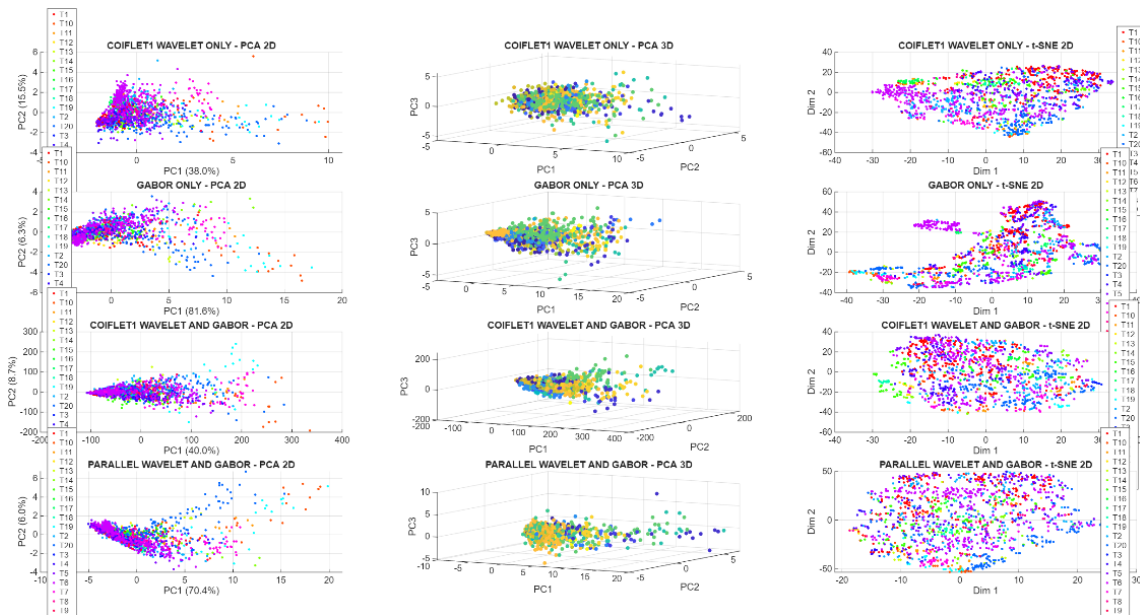


FIGURE 7. PCA and t-SNE visualizations of fusion strategies.

CONCLUSIONS

A comparative analysis of handcrafted feature to image transformation methods shows that both Coiflet1 Wavelet and Gabor representations have produced robust discriminative embeddings for turtle identity classification. The high accuracy, recall, and AUC values of this data imply that spatial frequency-based descriptors, when transformed to image like representations, retain class distinct information that can be used in convolutional processing. The Coiflet1 wavelet-based methods had the most consistent performance over all metrics evaluated with their effectiveness in representing structured and scale invariant scute facial structures. Similarly, Gabor representation showed strong separability, especially to represent orientation sensitive texture information. Sequential fusion of Coiflet1 with Gabor feature resulted in bad classification consistency and low feature compactness and appeared to be redundant due to a high number of frequency domain information overlapping. The parallel fusion approach attenuated some of this loss but failed to surpass the hand-crafted feature methods as a whole suggesting that naive-wise feature concatenation may decrease the discrimination strength rather than amplify it. Collectively, the results confirm that handcrafted feature transformations retain their appropriateness as interpretable and computationally inexpensive front-end representations in CNN-based individual turtle identifications. Even more importantly, it emphasises that effective incorporation of parallel handcrafted descriptors needs adaptive or learned fusion strategies to avoid redundancy as well as achieve maximum representation synergy. It gives us a good direction for later study towards hybrid representation learning architectures on fine-grained marine wildlife identification.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the Ministry of Higher Education Malaysia for providing funding through the Fundamental Research Grant Scheme (FRGS) under grant number FRGS/1/2020/ICT02/UMT/02/1 (Vote No. 59621) and the Fisheries Research Institute Rantau Abang for their valuable contributions to this research work.

REFERENCES

- Adam, L., Papafitsoros, K., Jean, C., Rees, A. F., & Čermák, V. 2025. Exploiting facial side similarities to improve AI-driven sea turtle photo-identification systems. *Ecological Informatics*, 103158.
- Abu, A., & Diamant, R. 2023. Underwater object classification combining SAS and transferred optical-to-SAS Imagery. *Pattern Recognition*, 144, 109868.
- Abdellatef, E., Omran, E. M., Soliman, R. F., Ismail, N. A., Abd Elrahman, S. E. S., Ismail, K. N., ... & Eisa, A. A. 2020. Fusion of deep-learned and hand-crafted features for cancelable recognition systems. *Soft Computing*, 24(20), 15189-15208.
- Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-Dabbagh, B. S. N., & Gu, Y. 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, 10(1), 46.
- Alenizy, H. A., & Berri, J. 2025. Transforming tabular data into images via enhanced spatial relationships for CNN processing. *Scientific Reports*, 15(1), 17004.
- Alshazly, H., Linse, C., Barth, E., & Martinetz, T. (2019). Handcrafted versus CNN features for ear recognition. *Symmetry*, 11(12), 1493.
- Artabaz, S., & Sliman, L. 2025. Feature fusion and selection using handcrafted vs. deep learning methods for multimodal hand biometric recognition. *Scientific Reports*, 15(1), 29237.
- Akbar, M. B. 2020. Wavelet analysis and comparison from coiflet family on image compression. In 2020 8th *International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-5). IEEE.
- Artabaz, S., & Sliman, L. 2025. Feature fusion and selection using handcrafted vs. deep learning methods for multimodal hand biometric recognition. *Scientific Reports*, 15(1), 29237.
- Bento, N., Rebelo, J., Barandas, M., Carreiro, A. V., Campagner, A., Cabitza, F., & Gamboa, H. 2022. Comparing handcrafted features and deep neural representations for domain generalization in human activity recognition. *Sensors*, 22(19), 7324.
- Bhati, D., Pachori, R. B., & Gadre, V. M. 2017. A novel approach for time–frequency localization of scaling functions and design of three-band biorthogonal linear phase wavelet filter banks. *Digital Signal Processing*, 69, 309-322.
- Can, Z., & Aydin, E. 2025. Explainable CNN–Radiomics Fusion and Ensemble Learning for Multimodal Lesion Classification in Dental Radiographs. *Diagnostics*, 15(16), 1997.
- Daugman, J. G. 1985. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7), 1160-1169.
- Daubechies, I. 1992. Ten lectures on wavelets. *Society for Industrial and Applied Mathematics*.
- Devulapalli, S., Potti, A., Krishnan, R., & Khan, M. S. 2023. Experimental evaluation of unsupervised image retrieval application using hybrid feature extraction by integrating deep learning and handcrafted techniques. *Materials Today: Proceedings*, 81, 983-988.
- De Zarzà, I., De Curtò, J., & Calafate, C. T. 2023. Optimizing neural networks for imbalanced data. *Electronics*, 12(12), 2674.
- Ding, J., Hu, J., Lin, J., & Zhang, X. 2024. Lightweight enhanced YOLOv8n underwater object detection network for low light environments. *Scientific Reports*, 14(1), 27922.
- Fujieda, S., Takayama, K., & Hachisuka, T. 2017. Wavelet convolutional neural networks for texture classification. *arXiv preprint arXiv:1707.07394*.
- Hays, G. C., Laloë, J. O., & Seminoff, J. A. 2025. Status, trends and conservation of global sea turtle populations. *Nature Reviews Biodiversity*, 1(2), 119-133.
- Hays, G. C., & Hawkes, L. A. 2018. Satellite tracking sea turtles: opportunities and challenges

- to address key questions. *Frontiers in Marine Science*, 5, 432.
- He, K., Zhang, X., Ren, S., & Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.
- Hu, J., Li, D., Duan, Q., Han, Y., Chen, G., & Si, X. 2012. Fish species classification by color, texture and multi-class support vector machine using computer vision. *Computers and Electronics in Agriculture*, 88, 133-140.
- Hu, K., Weng, C., Zhang, Y., Jin, J., & Xia, Q. 2022. An overview of underwater vision enhancement: From traditional methods to recent deep learning. *Journal of Marine Science and Engineering*, 10(2), 241.
- Jeantet, L., Zondo, K., Delvenne, C., Martin, J., Chevallier, D., & Dufourq, E. 2024. Automatic identification of the endangered hawksbill sea turtle behavior using deep learning and cross-species transfer learning. *Journal of Experimental Biology*, 227(24), jeb249232.
- Kamarainen, J. K., Kyrki, V., & Kalviainen, H. 2006. Invariance properties of Gabor filter-based features-overview and applications. *IEEE Transactions on image processing*, 15(5), 1088-1099.
- Lahmiri, S., & Boukadoum, M. 2013. Hybrid discrete wavelet transform and gabor filter banks processing for features extraction from biomedical images. *Journal of Medical Engineering*, 2013(1), 104684.
- LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. *nature*, 521(7553), 436-444.
- Lee, T. S. (1996). Image representation using 2D Gabor wavelets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10), 959-971.
- Lahiri, M., Tantipathananandh, C., Warungu, R., Rubenstein, D. I., & Berger-Wolf, T. Y. 2011. Biometric animal databases from field photographs: identification of individual zebra in the wild. In *Proceedings of The 1st ACM International Conference On Multimedia Retrieval* (pp. 1-8).
- Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., & Fisher, R. B. (2016, September). Coral classification with hybrid feature representations. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 519-523). IEEE.
- Mallat, S. G. 2002. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674-693.
- Nakka, R., Harursampath, D., & Ponnusami, S. A. 2023. A generalised deep learning-based surrogate model for homogenisation utilising material property encoding and physics-based bounds. *Scientific Reports*, 13(1), 9079.
- Neves-Ferreira, I., Mello-Fonseca, J., & Ferreira, C. E. 2023. Photo-identification shows the spatio-temporal distribution of two sea turtle species in a Brazilian developmental foraging ground. *Marine Biology*, 170(7), 83.
- Panarese, D., Alshehhi, M. A., Al Dhahouri, M. J., Sarrouf Willson, M., Köhnik, S., Stelfox, M., & Lloyd, J. R. 2025. Photo-ID for Capture Mark Recapture Study of Sea Turtles in Musandam, Oman, 2019–2023. *Chelonian Conservation and Biology: Celebrating 25 Years as the World's Turtle and Tortoise Journal*, 24(1), 111-124.
- Parmar, K., Kher, R. K., & Thakkar, F. N. 2012. Analysis of CT and MRI image fusion using wavelet transform. In *2012 International Conference on Communication Systems And Network Technologies* (pp. 124-127). IEEE.
- Riaz, F., Hassan, A., & Rehman, S. 2018. Texture classification framework using gabor filters and local binary patterns. In *Science and Information Conference* (pp. 569-580). Cham: Springer International Publishing.
- Santos, C. A. G., Freire, P. K. M. M., Silva, G. B. L., & Silva, R. M. 2014. Discrete wavelet transform coupled with ANN for daily discharge forecasting into Três Marias

- reservoir. *Proceedings of the International Association of Hydrological Sciences*, 364, 100-105.
- Schettini, R., & Corchs, S. 2010. Underwater image processing: state of the art of restoration and image enhancement methods. *EURASIP Journal On Advances In Signal Processing*, 2010(1), 746052.
- Stavelin, H., Rasheed, A., San, O., & Hestnes, A. J. 2021. Applying object detection to marine data and exploring explainability of a fully convolutional neural network using principal component analysis. *Ecological Informatics*, 62, 101269.
- Sharma, A., Vans, E., Shigemizu, D., Boroevich, K. A., & Tsunoda, T. 2019. DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific Reports*, 9(1), 11399.
- Simonyan, K., & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Tang, H., Li, Y., Huang, Z., Zhang, L., & Xie, W. 2022. Fusion of multidimensional CNN and handcrafted features for small-sample hyperspectral image classification. *Remote Sensing*, 14(15), 3796.
- Tabuki, K., Nishizawa, H., Abe, O., Okuyama, J., & Tanizaki, S. 2021. Utility of carapace images for long-term photographic identification of nesting green turtles. *Journal of Experimental Marine Biology and Ecology*, 545, 151632.
- Tschuchnig, M., Gadermayr, M., & Djemal, K. 2025. Hybrid Deep Learning and Handcrafted Feature Fusion for Mammographic Breast Cancer Classification. *arXiv preprint arXiv:2507.19843*.
- Ullah, W., Khalid, Y. N., & Khan, S. H. 2025. A Novel Deep Hybrid Framework with Ensemble-Based Feature Optimization for Robust Real-Time Human Activity Recognition. *arXiv preprint arXiv:2508.18695*.
- Vidal, M., Wolf, N., Rosenberg, B., Harris, B. P., & Mathis, A. 2021. Perspectives on individual animal identification from biology and computer vision. *Integrative and Comparative Biology*, 61(3), 900-916.
- Wallace, B. P., DiMatteo, A. D., Hurley, B. J., Finkbeiner, E. M., Bolten, A. B., Chaloupka, M. Y., ... & Mast, R. B. 2010. Regional management units for marine turtles: a novel framework for prioritizing conservation and research across multiple scales. *Plos one*, 5(12), e15465.
- Yang, L., Gu, Y., & Feng, H. 2025. Multi-scale feature fusion and feature calibration with edge information enhancement for remote sensing object detection. *Scientific Reports*, 15(1), 15371.
- Zhu, Y., Brettin, T., Xia, F., Partin, A., Shukla, M., Yoo, H., ... & Stevens, R. L. 2021. Converting tabular data into images for deep learning with convolutional neural networks. *Scientific reports*, 11(1), 11325.
- Zhao, F., Zhang, C., & Geng, B. 2024. Deep multimodal data fusion. *ACM Computing Surveys*, 56(9), 1-36.