Wadi Alshatti University Journal of Pure and Applied Sciences

مجلة جامعة وادي الشاطئ للعلوم البحتة والتطبيقية

Volume 4, No. 2, July-December 2026 Online ISSN: 3006-0877 المجلد 4، الاصدار 2، يوليو - ديسمبر 2026

RESEARCH ARTICLE COMMUNICATION ENGINEERING

Wavelet-Enhanced DEMON Processing and Deep Learning for Passive SONAR Target Recognition

Saaida Elmahdi Alnour [1,*], Marai Abousetta [1]

[1]Electrical and Computer Department,School of Engineering and Applied Sciences, Libyan Academy, Tripoli, Libya.

ARTICLE HISTORY

Received 04 July 2026
Revised 30 July 2026
Accepted 09 August 2026
Online 11 August 2026

KEYWORDS

Passive sonar;
Discrete wavelet transform;
DEMON;
Convolutional neural networks;
Underwater acoustics.

ABSTRACT

This paper presents an integrated underwater Automatic Target Recognition (ATR) framework that synergizes advanced signal processing with deep learning. To eliminate non-stationary ambient ocean noise and isolate low-frequency modulations, the raw acoustic pressure signals undergo pre-denoising via Discrete Wavelet Transform (DWT with db4 mother wavelet and soft-thresholding), followed by Detection of Envelope Modulation on Noise (DEMON) analysis to yield 128-bin spectral feature vectors. Classification is performed using a dedicated 7-layer 1D Convolutional Neural Network (1D-CNN) architecture terminating in a Softmax output layer. Acoustic simulations in MATLAB R2024b generated a balanced dataset of 1,200 audio samples (200 samples per class across six distinct targets: fast boat, rubber zodiac, human diver, giant turtle, dolphin, and mini submarine) propagated through a 20–40 path shallow-water multipath channel ($f_s = 10$ kHz). Experimental evaluations demonstrate an overall classification accuracy of 94.93% and a macro-averaged F1-score of 0.9505. Coupled with an average inference latency of 11.40 ms per target and Euclidean distance-based thresholding for out-of-distribution anomaly detection, the proposed system proves computationally efficient and well-suited for real-time passive underwater surveillance.

معالجة إشارات DEMON المعززة بتقنية المويجات (Wavelet) والتعلم العميق للتعرف على الأهداف باستخدام السونار السلبي

سعيدة المهدي النور [*,1]، مرعي ابوستة [1]

الكلمات المفتاحية:

السونار السلبي
تحويل المويجات المتقطع (DWT)
DEMON
الشبكات العصبية التلافيفية (CNN)
الصوتيات تحت الماء

الملخص

تقدم هذه الورقة إطارًا متكاملًا للتعرف الآلي على الأهداف تحت الماء (ATR) يدمج بين معالجة الإشارات المتقدمة والتعلم العميق. ولإزالة الضوضاء المحيطية غير المستقرة في المحيط وعزل التعديلات منخفضة التردد، تخضع إشارات الضغط الصوتي الخام لإزالة الضوضاء مسبقاً باستخدام تحويل المويجات المتقطع (DWT باستخدام المويجات الأم db4 والترشيح بالعتبة اللينة)، يليه تحليل كشف تعديل الغلاف على الضوضاء (DEMON) لإنتاج متجهات خصائص طيفية بأبعاد 128 خانة. وتتم عملية التصنيف باستخدام بنية شبكة عصبية تلافيفية أحادية الأبعاد (1D-CNN) مخصصة مكونة من 7 طبقات تنتهي بطبقة مخرجات بدالة Softmax. ولتقييم النموذج، تم توليد مجموعة بيانات متوازنة عبر المحاكاة الصوتية في برنامج MATLAB R2024b تتكون من 1,200 عينة صوتية (200 عينة لكل فئة عبر ستة أهداف مختلفة: قارب سريع، قارب مطاطي، غواص بشري، سلحفاة ضخمة، دولفين، وغواصة صغيرة) تنتشر عبر قناة متعددة المسارات في المياه الضحلة تتراوح بين 20 و40 مساراً بتردد أخذ عينات $f_s = 10$ kHz. وتُظهر التقييمات التجريبية دقة تصنيف إجمالية تبلغ 94.93% ومعامل F1 بمتوسط كلي يبلغ 0.9505. وإلى جانب زمن استجابة استنتاج يبلغ 11.40 مللي ثانية لكل هدف واعتماد آلية عتبة قائمة على المسافة الإقليدية لكشف الشذوذ للأهداف خارج التوزيع، يثبت النظام المقترح كفاءته الحسابية وملاءمته العالية للمراقبة البحرية السلبية تحت الماء في الوقت الفعلي.

Introduction

Sonar systems, which are often classified into active and passive sonar systems based on their method of operation, are widely utilized for underwater detection, identification, localization, navigation, and communication, in contrast to active sonar (AS), passive sonar (PS) relies solely on the acoustic signals it receives to assess the state of the ocean environment, identify possible targets, and initiate the appropriate signal processing, spatial signal processing using receiving arrays is central in modern acoustic systems that can significantly improve the quality of signal pickup in reverberant and noisy environments [1], in the complex ocean environment governed by the physics of computational ocean acoustics [2], these arrays are increasingly utilized to extract precise acoustic features of marine targets by measuring signatures using the DEMON algorithm in passive sonar [3], with the rapid advancement of artificial intelligence, passive tracking techniques that integrate spectral analysis and deep learning have emerged as important tools [4].

Recent survey studies focused on assessing the role of machine learning in analyzing ship noise [5]. An example of this is the study, which applied DEMON-type algorithms in ship and diver detection [6]. In this, it was found that high

*Corresponding author
https://doi.org/10.63318/waujpasv4i2_26

performance is achieved if the systems are built with high-performance Convolutional Neural Networks (CNNs) [7], and these frameworks increased performance with integrated signal denoising and hybrid modeling[8]. The structure of the acoustic sensor is based on the sequential application of simple noise reduction algorithms [9]. This structure has proven very effective in improving acoustic sensor performance.

Underwater beam forming is a method used to spatially steer a network of hydrophones to achieve signal isolation and enhancement [10]. The combination of AI and the signal preprocessing techniques above have made it possible to formulate networks such as DEMONet [11], and Mixed-Experts models for target recognition and localization [12]. This method expands the potential of sonar systems to isolate the specific noise patterns of small crafts and other targets [13].

Related Work

Researchers use advanced physical models to understand the complex signal interferences in water. One of these models is the waveguide invariant theory [14]. This theory is important for the optimal geometric placement of passive sonar systems for the best coverage in underwater surveillance [15]. Building on these models, there was a shift to using a Dense Convolutional Neural Network (Dense CNN) to analyze acoustic data and recognize targets in the marine environment [16]. Deep learning for target detection underwater, and in particular for the integration of neural networks with classical spectral analysis, has been reviewed in detail, and is extremely promising [17]. The integration of neural networks and spatial beamforming has proven incredibly successful in shallow water acoustic engagement [20]. The integration of machine learning in neural networks has also proven successful [21]. This trend has been documented in many surveys that have tracked the methods of underwater target recognition [22]. By focusing on the most relevant spectral attributes, Attention-based DNN networks help achieve higher classification accuracy [23]. In order to accomplish this, the acoustic signal needs to be cleaned of background noise. This process has a strong mathematical basis, at the core of this complex processing pipeline lies the urgent need to purify the acoustic signal from random background noise before feeding it into intelligent classification networks, this denoising process is mathematically rooted in soft-thresholding algorithms [24], which represent the cornerstone of the wavelet transform filters utilized in the methodology proposed in this paper.

Problem Statement

Continuous and reliable monitoring of the marine environment is necessary to effectively protect offshore assets; therefore, a critical means of monitoring and identifying a variety of marine targets, such as surface and submerged vessels, and to employ passive sonar systems for identification of all divers. The shallow-water acoustic channel provides a very difficult and complex environment for precise signal processing because of the significant degradation of acoustic signatures acquired via hydrophone arrays due to multipath, heavy ambient ocean noise and dynamic interference, resulting in low SNR. Standard automated target recognition and signal filtering techniques typically do not work well under these extreme conditions. There exist significant challenges inherent in the use of conventional filters to remove non-linear, non-stationary environmental noise while simultaneously preserving low frequency target features such as cavitation and modulation frequencies.

In addition, standard feature extraction techniques are consistently producing distorted or inaccurate signatures which result in reduced accuracy or robustness of the classification techniques based on those features. This research will provide solutions to address the significant vulnerability of current passive sonar classification systems to extreme acoustic degradation and the urgent need for an automated framework to effectively denoise raw hydrophone data, maintain the physical modulations of the target, and utilize state-of-the-art deep learning architectures to provide high confidence classification in realistic, high noise level, underwater environments.

Methodology

This study offers an all-encompassing structure for the application of automatic target recognition technology in marine settings by utilizing advanced signal processing methods, along with a deep learning model, that can successfully operate in extremely noisy underwater environments. The methodology consists of several stages, including:

1. Data modeling and simulation of the acoustic environment—to evaluate and train the system, passive acoustic assertions were simulated for six types of marine target classes (i.e. fast boats, submarines, and human divers), including multipath channel effects reflecting surfaces/dispersal effects due to distance from source(s); the use of pink noise added to background sound at low signal-to-noise ratio (SNR) provided a simulation for the adverse effect of ocean conditions; and array beam forming to improve the spatial gain of the target's audio signal before processing.

2. Wavelet-Based Denoising and DEMON Feature Extraction.

To suppress non-stationary marine ambient noise and dynamic ocean multipath interference prior to envelope demodulation, the acoustic pressure time-series signals undergo pre-denoising via Discrete Wavelet Transform (DWT). Multi-resolution decomposition is executed using the Daubechies 4 (db4) mother wavelet across 4 decomposition levels. Universal soft-thresholding is applied to the detail coefficients at each level to filter out high-frequency thermal and ambient fluctuations while preserving essential structural transients.

3. Following wavelet denoising, Detection of Envelope Modulation on Noise (DEMON) analysis is performed by high-pass filtering, full-wave rectification, low-pass filtering, and Fast Fourier Transform (FFT) processing to extract 128-bin normalized spectral feature vectors (128×1) representing low-frequency shaft and blade passage modulations.

4. **1D-CNN Network Topology**

The classification pipeline relies on a lightweight, 7-layer 1D Convolutional Neural Network (1D-CNN) optimized for spectral pattern recognition. As detailed below, the 7 operational layers are structured as follows:

1. Input Layer: Accepts normalized 128×1 DWT-DEMON spectral feature vectors.
2. Convolutional Layer 1 (Conv Block 1): Consists of 20 filters with a 1×7 kernel size, applied with a stride of 1 and 'same' zero-padding to preserve boundary spectral energy. This layer is integrated with Batch Normalization (BN) and Rectified Linear Unit (ReLU)

activation.

3. Max Pooling Layer 1: Applies a 1×2 pool size with a stride of 2 to downsample feature map dimensions to 64×20.
4. Convolutional Layer 2 (Conv Block 2): Expands depth using 40 filters with a 1×5 kernel size, utilizing a stride of 1 and 'same' padding, coupled with BN and ReLU.
5. Max Pooling Layer 2: Uses a 1×2 pool size with a stride of 2 to condense the latent output down to 32×40.
6. Fully Connected Dense Layer (FC1): Flattens features into an 80-neuron layer with ReLU activation, regulated by a Dropout rate of 0.35 to prevent model overfitting.
7. Output Layer (FC2): Comprises 6 neurons with a Softmax activation function to compute class probability distributions across the target categories.

The spectral data was processed by a custom one-dimensional convolutional neural network (1D CNN) to classify the target recognition task using (1D-CNN). The 1D-CNN consists of several Convolutional Layers operating in succession, all of which are supported by Batch Normalization for stability during training and also have Max Pooling layers to reduce dimensionality, a set of fully connected hidden layers with dropout mechanisms has been added to the model in order to avoid overfitting.

Mathematical Modelling

To formally evaluate the proposed hybrid work, the physical acoustic environment and the signal processing stages are mathematically formulated, the system model encompasses the multipath propagation channel, array beamforming, envelope extraction, and the wavelet transform formulation.

A. Underwater Acoustic Propagation Model

In a shallow-water offshore environment, the acoustic signal emitted by a target undergoes multipath propagation. Let $s(t)$ represent the broadband source signal modulated by the propeller blade rate. The received acoustic signal at the m-th hydrophone, denoted as $x\_m(t)$, is modeled as:

$$x_m(t) = \sum_{p=1}^{p} A_{m,p} s(t - \tau_{m,p}) + \eta_m(t) \qquad (1)$$

where $A_{m,p}$ is the attenuation factor for the p-th path, $\tau_{m,p}$ is the corresponding time delay, and $\eta_m(t)$ represents the additive ambient marine noise at the m-th sensor.

B. Array Spatial Filtering

For a Uniform Linear Array (ULA) of M sensors with inter-element spacing d , Delay-and-Sum Beamforming is employed. The beamformer output $y(t)$ steered towards the target's azimuth angle θ is defined as:

$$y(t) = \frac{1}{M} \sum_{m=1}^{M} w_m x_m(t + \Delta_m(\theta)) \qquad (2)$$

where w_m is the amplitude weighting for the $m-th$ sensor, and $\Delta_m(\theta)$ is the time delay applied to align the signals coherently, given by:

$$\Delta_m(\theta) = \frac{(m-1)d\sin(\theta)}{c} \qquad (3)$$

Where c is the speed of sound in water.

Conventional beamforming (CBF), which uses the phase of the signals received by a particular array element as a reference, is the most popular beamforming technique, (CBF) transforms the signal from the temporal domain to the spatial domain by a procedure called "phase alignment and accumulation" [1]. As illustrated in Figure 1, delay-and-sum beamforming (DASB), one of the basic types of spatial filtering, entails delaying the outputs from several sensors by predetermined amounts in order to align the spatial components of the signals arriving from the target's direction, followed by summation. When the beamformer is directed toward a single target, this technique maximizes the average output power. By using a hydrophone array to steer the "listening" direction towards the target, thereby suppressing noise from other directions and increasing the SNR (Array Gain), the equation illustrated in formula (2), and as shown in figure (1).

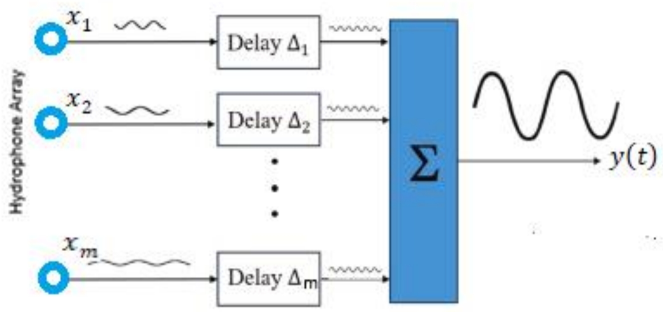

Figure 1: Beamforming Dalay-and -Sum

C. Acoustic Signal Denoising via Wavelet Transform

Prior to getting the modulation envelope, the raw beamformed acoustic signal $x_b(t)$ is affected by transient hydrodynamic interference and non-stationary cavitation noise. It is necessary to reduce this background noise while maintaining the time characteristics of the target's modulation. Therefore, direct implementation of the 4-level Discrete Wavelet Transform (DWT) of the signal $x_b(t)$ is applied with the use of the Daubechies 4 (db4) wavelet. While carrying out the Discrete Wavelet Transform the signal $x_b(t)$, is processed in such a way that it decomposes into approximation coefficients cA_J and detailed coefficients cD_j while being decomposed up to a certain degree J=4:

$$x_b(t) = \sum_k c\, A_J[k]\, \phi_{J,k}(t) + \sum_{j=1}^{J} \sum_k c\, D_j[k]\, \psi_{j,k}(t) \qquad (4)$$

where $\phi_{J,k}(t)$ and $\psi_{j,k}(t)$ denote the scaling and mother wavelet functions, respectively.

To eliminate non-stationary noise components, the noise standard deviation $\hat{\sigma}$ is estimated dynamically from the first-level detail coefficients cD_1 using Donoho's Median Absolute Deviation (MAD) scale estimator:

$$\hat{\sigma} = \frac{\mathrm{median}(|cD_1|)}{0.6745} \qquad (5)$$

Using $\hat{\sigma}$, a universal soft-thresholding rule is applied across all detail coefficients cD_j:

$$\lambda = \hat{\sigma}\sqrt{2\ln(N)} \qquad (6)$$

$$\widetilde{cD}_j[k] = \mathrm{sgn}(cD_j[k]) \cdot \max(0, |cD_j[k]| - \lambda) \qquad (7)$$

where N represents the length of the acoustic signal frame ($N = 7{,}000$), λ is the universal threshold value, and $\widetilde{cD}_j$ denotes the thresholded detail coefficients.

The denoised acoustic signal $\hat{x}_b(t)$ is subsequently reconstructed by applying the Inverse Discrete Wavelet

Transform (IDWT) using the preserved approximation coefficients cA_4 and the thresholded detail coefficients $\widetilde{cD}_j$:

$$\hat{y}(t) = \mathrm{IDWT}\left(cA_4, \widetilde{cD}_1, \widetilde{cD}_2, \widetilde{cD}_3, \widetilde{cD}_4\right) \quad (8)$$

This stabilized, high-SNR signal $\hat{x}_b(t)$ serves as the clean input for the subsequent DEMON processing block.

D. DEMON Envelope Extraction

With the broadband noise suppressed, the DEMON algorithm is applied to the denoised signal $\hat{y}(t)$ to isolate the amplitude modulations caused by the target's propeller, first, the analytic signal $z(t)$ is constructed using the Hilbert Transform $\mathcal{H}\{\cdot\}$ to remove the negative frequency components:

$$z(t) = \hat{y}(t) + j\mathcal{H}\{\hat{y}(t)\} \quad (9)$$

where the Hilbert Transform of the denoised signal is defined as:

$$\mathcal{H}\{\hat{y}(t)\} = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\hat{y}(\tau)}{t-\tau} d\tau \quad (10)$$

the instantaneous modulation envelope $E(t)$ is computed as the magnitude of the analytic signal:

$$E(t) = |z(t)| = \sqrt{\hat{y}^2(t) + \mathcal{H}^2\{\hat{y}(t)\}} \quad (11)$$

This robust envelope $E(t)$ is subsequently transformed into the frequency domain via STFT to generate the clear spectrograms required for the CNN classifier.

E. Spectrogram Generation (STFT)

Convolutional Neural Networks (CNNs) process two-dimensional data highly effectively. Therefore, the stabilized 1D temporal envelope $E(t)$ must be transformed into a 2D time-frequency representation using the Short-Time Fourier Transform (STFT).

The discrete spectrogram matrix $S(m, k)$ at time frame m and frequency bin k is computed as:

$$S(m,k) = |\sum_{n=0}^{N-1} E[n+mH]w[n]e^{-j\frac{2\pi kn}{N}}|^2 \quad (12)$$

Where: $w[n]$ is the window function (e.g., Hamming window) used to mitigate spectral leakage, N is the window length, and H is the hop size.

The resulting matrix $S(m, k)$ represents the DEMON spectrogram image, which serves as the input feature map for the deep learning model.

F. Feature Extraction via CNN

The modified CNN architecture relies on depthwise separable convolutions to reduce computational complexity while maintaining high feature extraction accuracy [26], a single spatial filter is applied to each input channel independently to extract local acoustic features (e.g., blade rate harmonics), for an input feature map F and a filter kernel K, the output G is given by:

$$G_{x,y,c} = \sum_{i=1}^{D_k}\sum_{j=1}^{D_k} K_{i,j,c} \cdot F_{x+i-1,y+j-1,c} \quad (13)$$

where D_k is the spatial dimension of the kernel, and c is the channel index 2, pointwise convolution equal to $A\ 1 \times 1$ convolution is then applied to linearly combine the outputs of the depthwise layer across all channels. The final output P is computed as:

$$P_{x,y,n} = \sum_{c=1}^{C_{in}} W_{c,n} \cdot G_{x,y,c} \quad (14)$$

where W represents the weights of the pointwise filter, C_{in} is the number of input channels, and n is the index of the output filter [27].

Following each convolutional operation, a non-linear Rectified Linear Unit (ReLU) activation function is applied along with Batch Normalization (BN) to accelerate convergence and stabilize training:

$$A_{x,y,n} = \max\left(0, BN\left(P_{x,y,n}\right)\right) \quad (15)$$

G. Classification Head

At the final stages of the network, to avoid the heavy computational burden of Fully Connected (FC) layers, the high-dimensional feature maps are passed through a Global Average Pooling (GAP) layer.

The GAP layer flattens the 2D spatial dimensions into a 1D vector v, where each element v_c is calculated as the spatial average of the corresponding feature map:

$$v_c = \frac{1}{H \times W}\sum_{x=1}^{H}\sum_{y=1}^{W} A_{x,y,c} \quad (16)$$

where H and W denote the height and width of the final feature map.

The vector v is fed into a Softmax classifier to compute the predicted probability distribution over the target classes (e.g., human divers, mini-submarines, etc.). The probability $\hat{P}_i$or class i is calculated using:

$$\hat{P}_i = \frac{e^{vi}}{\sum_{i=1}^{C} e^{vj}} \quad (17)$$

where C is the total number of target classes (in this study, $C = 6$). The class with the maximum probability $\hat{P}$ commands the network's final target recognition decision.

Results and Discussion

This section offers a thorough analysis of the suggested hybrid approach for underwater acoustic signal processing and classification; in order to prove the reliability and effectiveness of the system, a number of tests have been carried out to analyze the different stages of the processing chain starting with the enhancement of signals spatially and temporally until the AI-based classification of the target signal.

To evaluate the proposed sonar recognition framework under realistic shallow-water operational conditions, a synthetic acoustic dataset comprising six target classes was developed: Fast Boat, Rubber Zodiac, Human Diver, Mini Submarine, Dolphin, and Large Turtle.

Acoustic signatures were synthesized at a sampling frequency of $f_s = 10$ kHz over a frame duration of $T = 0.7$ s , corresponding to $N = 7{,}000$ discrete time-domain samples per acoustic realization. Spatial array processing gain was incorporated by modeling an 8-element Uniform Linear Array (ULA). Acoustic propagation through the ocean medium was modeled using a discrete multipath channel with 20 to 40 random paths and exponential decay:

Feature Analysis: Spectral integrity and Correlation:

Before using the data as an input to the CNN, it is essential to make sure that the extracted features are actually physical kinematics and not random noise. Figure 2, shows the Pearson Correlation heatmap of the extracted DEMON spectral features. In order to ensure the effectiveness of the CNN and minimize the risk of over-fitting, a Pearson Correlation analysis was applied to the extracted feature vector. The entire input vector is N = 128 frequency bins long, which corresponds to the whole spectrum of DEMON used for target classification. The calculation of the correlation matrix was carried out in full dimensions (128).

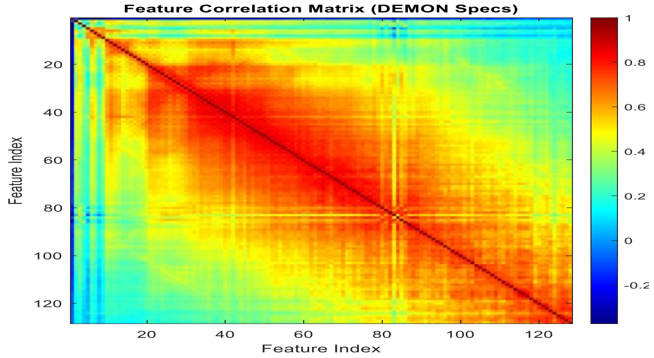

Figure 2: Pearson correlation heatmap for the extracted DEMON spectral features

For visualization purposes and highlighting the most physically significant phenomena, Figure 3 presents a magnified sub-matrix corresponding to only the first 16 frequency bins (Low-Frequency Region). This specific frequency range was chosen for the following reasons:
This region (approximately 0 Hz to 22 Hz) includes the main propeller shaft frequencies and their harmonics which are the most distinctive features for maritime objects, and considering this specific part enables a visual representation of "Harmonic Coupling" effects which will be invisible in a full-size (128 times) map.

Figure 3: The sequential time-domain signal processing

Figure 4 shows the time domain processing procedure of signals from one hydrophone (sequential processing pipeline) for feature extraction from underwater acoustic signal. The procedure starts with (1) receiving a raw noisy signal with a low SNR. After that (2) enhanced signal is obtained using Delay-and-Sum beamforming and Wavelet Transform denoising which keeps important cavitation bursts. Then (3) instantaneous signal envelope is extracted through DEMON processing which removes DC component and applies low pass filtering to obtain low frequency mechanical characteristic such as propeller blade rate.

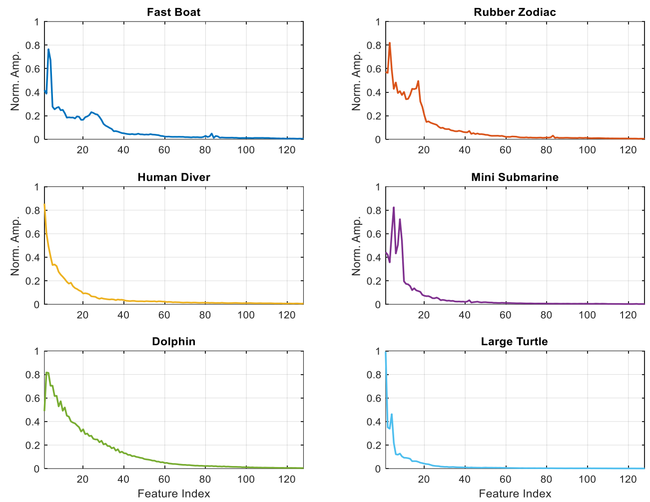

Figure 4: Average DEMON spectral signatures per target class

Figure 5, presents the selectivity and efficacy of the DWT preprocessing stage using a db4 wavelet with soft thresholding.

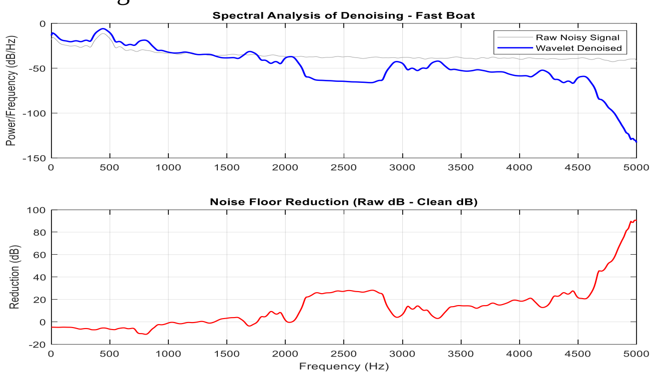

Figure 5: Wavelet denoising performance in the frequency domain for a fast boat sample

A - Spectral Power Density compares the raw noisy signal (grey) and the denoised signal (blue). The denoising is frequency-selective. In the region below 1,000 Hz, the blue line essentially tracks the grey line, thus demonstrating the preservation of the crucial low-frequency components of the target signature that are needed to carry out the DEMON analysis. In the bandwidth above 2,000 Hz, the blue line is well below the grey noise floor, indicating the suppression of broadband pink noise, which contains the dominant noise, and which contains the denoised signal.
B - Noise Floor Reduction shows (Raw - Clean dB). This displays significant noise reduction, with maximum noise reductions exceeding 80 dB at the highest frequencies. As a result, a specific sample's SNR is improved on the order of nearly 4 dB. This improves classification accuracy of the CNN in a safe and sane manner without signal corruption on lower frequencies.
Efficiency of the Wavelet Denoising Method: It was observed that the proposed preprocessing of the signal using the DWT algorithm with a $db4$ mother wavelet and soft thresholding has proven to be efficient against complex noise conditions in underwater environments.
The Root Mean Square Error ($RMSE$) of the pristine target signatures compared to the wavelet-denoised signals was obtained to be 4.0195 as given in equation

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x(n) - \hat{x}(n))^2} \qquad (18)$$

where $\boldsymbol{N}$ is the number of discrete samples of the acoustic signal, $x(n)$ is the amplitude of the pristine, noise-free target signature at the nth discrete point, and $\hat{x}(n)$ is the amplitude of the reconstructed signal after the discrete wavelet transform (DWT) denoising procedure at the nth discrete point.

This low error margin proves quantitatively that the wavelet method successfully reduced the broadband pink noise and multipath interference without heavily compromising the essential low-frequency modulations of the signals.

CNN:

Figure 6, shows the loss curves of training and validation, where there is an asymptotic and smooth convergence of both curves, as there is a decrease in training loss in the early epochs and remains stable without any oscillation. The fact that the training and validation accuracy curves are very close to each other indicates that there is no overfitting.

The 1D-CNN architecture proposed is presented in Table 1. The pipeline starts with inputs of normalized 128×1 DWT-DEMON spectral feature vectors passing through the first convolutional block with 20 filter units with a kernel of size 1×7, with Batch Normalization (BN) and Rectified Linear Unit activation. After that, a Max Pooling layer (pool size 2, stride 2) reduces the spatial dimensions to 64×20. The second convolution block then broadens the depth of the features by using 40 filters with a kernel of size 1×5, which again streamlines the latent representation to be of size 32×40 following the pooling. Finally, the feature maps pass through the fully connected dense layer (FC1) of 80 neurons with a dropout rate of 0.35 to reduce overfitting, finishing with an output layer (FC2) of 6 neurons with Softmax activation to produce the probabilities of classes across the six categories.

Table 1: Detailed Specifications of the Proposed 1D-CNN Architecture

Layer / Stage	Output Dimensions	Parameters & Configurations
Input Layer	$1 \times 128 \times 1$	Normalized DWT-DEMON features (128 bins)
Conv Block 1	$1 \times 128 \times 20$	20 filters, Kernel size: 1×7, Padding: 'same' + BN + ReLU
Max Pooling 1	$1 \times 64 \times 20$	Pool size: 1×2, Stride: 1×2
Conv Block 2	$1 \times 64 \times 40$	40 filters, Kernel size: 1×5, Padding: 'same' + BN + ReLU
Max Pooling 2	$1 \times 32 \times 40$	Pool size: 1×2, Stride: 1×2
Dense (FC1) + Dropout	80	80 Neurons + ReLU, Dropout Rate = 0.35
Output (FC2)	6	6 Target Classes + Softmax Layer

As per the hyperparameter settings and training methodology given in Table 2, the network uses a fixed learning rate schedule of $\eta = 0.001$ and a mini-batch size of 32 samples per iteration. The training is conducted over 20 epochs for a total of 700 gradient update cycles (35 iterations per epoch). Categorical Cross-Entropy is used as the multi-class loss function.

Table 2: Hyperparameters for Model Training

Hyperparameter	Value / Setting	Description
Initial Learning Rate	0.001 (1×10^{-3})	Constant learning rate schedule
Mini-Batch Size	32 samples	Training subset size per step
Maximum Epochs	20 epochs	Total passes through dataset
Total Iterations	700 steps	35 iterations per epoch
Loss Function	Categorical Cross-Entropy	Multi-class optimization loss

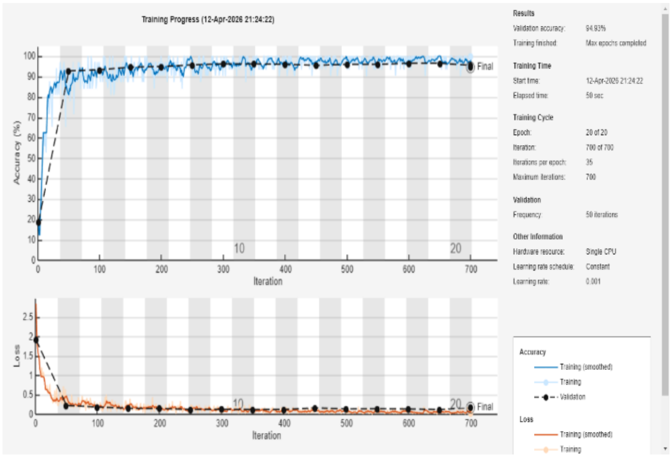

Figure 6: Accuracy and loss curve for training progress CNN

Figure 7 presents the confusion matrix for the developed algorithm where the True Labels are presented by the vertical axis and the Predicted Labels – by the horizontal one.

Diagonal Dominance - the matrix demonstrates the high concentration of the values along the diagonal, that proves the high accuracy of the samples classification, what is consistent with the high accuracy of the algorithm as a whole.

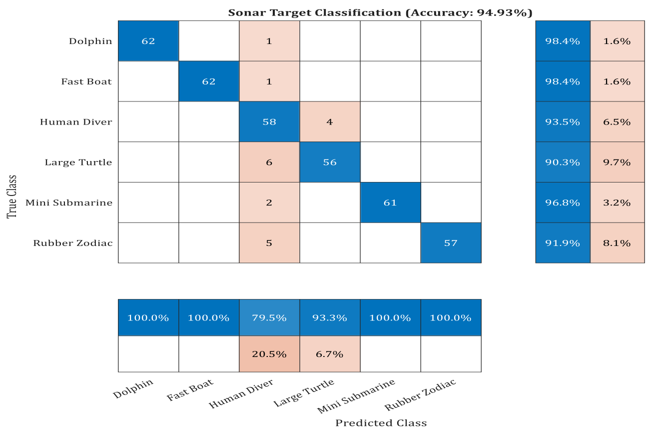

Figure 7: Confusion matrix showing the classification performance on the test set

As depicted in the Confusion Matrix, the proposed framework achieves high class precision (100.0%) for both the Fast Boat and Dolphin categories, with an overall recall of 98.4%. Specifically, out of 63 test samples per class, exactly 62 samples were correctly identified, while 1 sample from each class was misclassified into the Human Diver category (row 1, col 3 = 1; row 2, col 3 = 1). The high separability of these two classes stems from their distinct physical acoustics:

Fast Boat Dynamics: Fast surface vessels generate high-speed propeller rotation, resulting in intense tip-vortex cavitation and strong, continuous Blade Passage Frequency (BPF) harmonic modulations. These high-amplitude periodic envelopes yield highly pronounced, multi-harmonic spectral peaks in the DWT-DEMON profile that are orthogonal to low-speed acoustic targets.

Dolphin Biological Acoustics: Dolphin signatures consist strictly of high-frequency bio-sonar whistles and echolocation pulses ($> 1\ \mathrm{kHz}$) lacking any rotational mechanical components or shaft modulations. The complete absence of mechanical cavitation dynamics creates a highly distinct spectral profile compared to artificial vessels.

Nevertheless, the conducted analysis shows that the latent representations belonging to the Fast Boat and Dolphin signatures belong to separate clusters which have big Euclidean distance.

Physical Analysis of the Human Diver Acoustic Signature: A detailed class-wise performance evaluation indicates that while motorized vessels (fast boat, mini submarine, and rubber zodiac*)* exhibit highly distinct spectral peaks in their DEMON envelopes, the human diver class displays a comparatively lower classification confidence. From an acoustic physics perspective, this phenomenon is attributed to the fundamental mechanism of sound generation:

1. Lack of Mechanical Propeller Cavitation: Motorboats produce strong periodic amplitude fluctuations resulting from spinning propeller blades and rotates shaft. DEMON detects these harmonic carrier fluctuations well. In contrast, the noise made by a human diver comes mainly from scuba regulator demand valves, periodic breathing bubble collapse and hydrodynamic agitation due to movement of fins.

2. Spectral Similarity to Ambient Noise: The emission of the noise of open-circuit breathing devices does not have fixed frequency of rotational modulations. Rather, the noise of air bubbles bursting is presented in the form of transient broadband bursts substantially overlapping with changing ocean sounds, i.e. sea waves break and background thermal noise. By applying the 4-level db4 Discrete Wavelet Transform (DWT), high-frequency transient interference is suppressed, allowing the custom 1D-CNN to detect subtle low-frequency respiratory rhythms ($0.8\ \mathrm{Hz} - 1.2\ \mathrm{Hz}$) that would otherwise be completely masked in uncleaned low-SNR passive sonar signals.

As shown in Table 3, the proposed 1D CNN provided excellent classification performance on the unseen test data set with an accuracy of 94.93%.

The further breakdown of the macro-averaged measures into Precision, Recall, and F1-Score provides the numbers of 0.9546, 0.9491 , and 0.9505 , respectively, indicating a highly balanced classification model without any bias. Class-specific analysis confirms the excellent performance of the system in the detection of mechanical targets. Such differences are expected and realistic considering the properties of the underwater acoustics and are mainly caused by the fact that a human diver has a very weak acoustic signature and his breathing and other movements may be similar to transient noise. However, from the security point of view, the Recall rate of 0.9355 for this class is highly important as it guarantees a very low false negative rate despite the occasional false positive.

Table 3: Class-wise Performance Metrics of the Proposed 1D-CNN Classifier

CLASS NAME	PRECISION	RECALL	F1-SCORE
Fast Boat	1.0000	0.9841	0.9920
Rubber Zodiac	1.0000	0.9841	0.9920
Human Diver	0.7945	0.9355	0.8593
Mini Submarine	0.9333	0.9032	0.9180
Dolphin	1.0000	0.9683	0.9839
Large Turtle	1.0000	0.9194	0.9580
AVERAGE (MACRO)	0.9546	0.9491	0.9505

From the data shown in Table 4, quantifies the reliability of the anomaly detection system, the decision boundary (Anomaly Threshold) was calculated at a distance of 22.4776, safely beyond the Normal Sample Mean Distance of 9.9489, in order to keep known objects close and avoid false positives. Upon exposure of an acoustic signature not trained on the system, the object was mapped to the feature space at a very distant location of 70.0901. As the distance is very far beyond the defined threshold ($70.0901 > 22.4776$), the system accurately recognized the object as Unknown Anomaly. Surprisingly, the system classified the unknown anomaly as being the nearest to the 'Human Diver', indicating that the neural network has rationally tried to find the most acoustically similar object among the known classes before identifying the anomaly.

Table 4: Distance Metrics for Unknown Target (Anomaly) Detection in the Reduced Feature Space

Metric	Distance Value	Notes
Anomaly Threshold	22.4776	Decision boundary
Normal Sample Mean Distance	9.9489	Average distance for known targets
Anomaly Object Distance	70.0901	Exceeds threshold *(Distance to nearest class: Human Diver)*

Following on from the above two-dimensional approach, Figure 8 shows a more holistic view of the 128-dimensional deep feature space obtained from the 1D-CNN through a three-dimensional projection. The use of the first three Principal Components (PC1, PC2 and PC3) ensures that more of the underlying variance in the dataset is captured.

The distinct clustering of the data further reinforces the capacity of the neural network for spatial isolation of the six known acoustic marine classes.

As shown in the three-dimensional depiction, the anomaly decision boundary is now presented through a Threshold Sphere wireframe. This surrounds the global centroid point of operation (black 'X'), while the unknown target is depicted using the red star along with its respective distance vector (red line). It is clearly evident from the figure that the anomaly exists well outside the threshold sphere in multidimensional space. This makes for a very convincing illustration of the effectiveness of the framework in isolating new acoustic objects from trained ones.

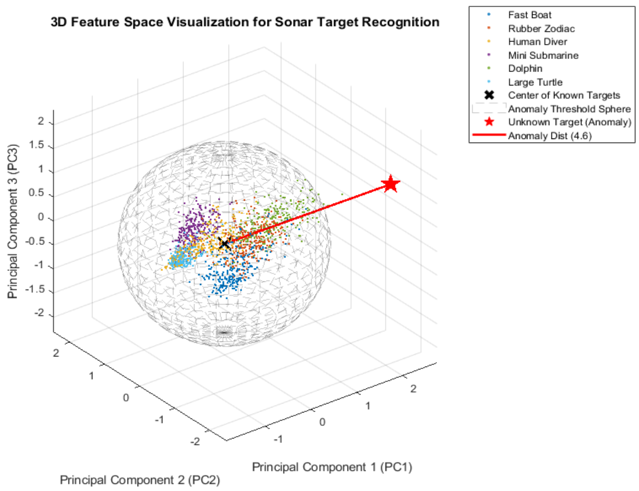

Figure 8: 3D Visualization of the PCA-reduced deep feature space and spherical anomaly decision boundary

Figure 9 illustrates ROC curves of the six classes of acoustic targets in assessing the capability of the 1D-CNN classifier through various discrimination thresholds. In most cases, the curves of nearly all mechanical and biological targets tend to approach very close to the top left corner, suggesting that the True Positive Rate (Sensitivity) is very high while at the same time the False Positive Rate is approaching zero. The area under the curve (AUC) metric supports this strong performance, where classes like 'Fast Boat' and 'Dolphin' achieve almost perfect AUC values of 0.9991 and 0.9960, respectively. Although there is slightly lower ROC curve in the case of 'Human Diver' class, the AUC value of 0.9760 is still excellent. This small difference indeed captures the physical difficulty in extracting the acoustic signature of the diver from the background noise of the ocean.

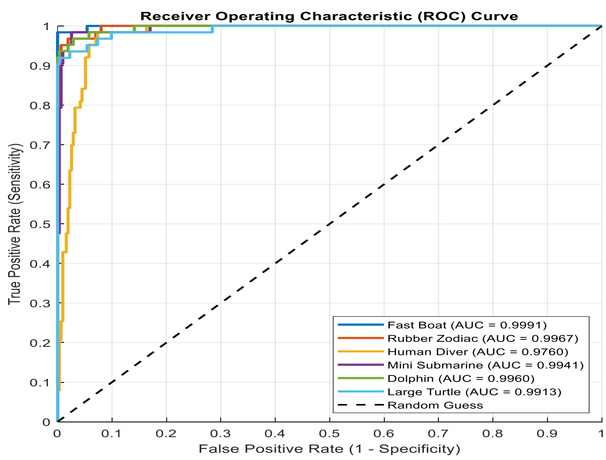

Figure 9: Receiver operating characteristic curves and AUC scores for the 1D-CNN multi-class classifier

The details relating to the important aspects of the modern research on underwater sound-based target identification are given in Table 5. The information includes target types, environmental settings, sampling frequency, classification efficacy, and multi-path dynamics. The research presented in this paper shows high-performance level by identifying 6 types of targets with efficiency of 94.93% accuracy using the methods of multi-path modelling for shallow water.

As shown in table 6; baseline classifier performance benchmarking, to evaluate the classification performance of the proposed deep 1D-CNN architecture, its performance was benchmarked against standard machine learning classifiers,

Table 5: Comparison with State-of-the-Art

Study	Target Classes	Envi.	Sampling Freq. (fs)	Accuracy
Jiang et al. 2022	General UATs	Shallow water	N/A	N/A
Jamal et al. 2024	Ships + Submarines	Noisy marine	>40 kHz (0–1 kHz used)	**93.75%**
Xie et al. 2024	Multiple ships (DeepShip, DTIL, ShipsEar)	Complex marine, unpredictable channels	N/A	(N/A)
Proposed Work	**6 Classes (Fast Boat, Zodiac, Submarine, Diver, Dolphin, Turtle)**	**Shallow water**	**10 kHz**	**94.93%**

specifically k-Nearest Neighbors ($k = 5$) and a Multiclass Support Vector Machine (SVM). All models were trained and evaluated using the identical 4-level db4 Discrete Wavelet Transform (DWT) and DEMON feature representation dataset ($N = 1{,}200$ acoustic signatures across 6 target classes).

Table 6: Benchmark Comparison of Classifiers on the DWT-DEMON Feature Matrix

Classifier Model	Preprocessing & Feature Input	Classification Accuracy (%)	Average Inference Latency (ms)
k-Nearest Neighbors ($k = 5$)	DWT (db4, Level 4) + DEMON (128 bins)	87.47%	10.05 ms
Support Vector Machine (SVM)	DWT (db4, Level 4) + DEMON (128 bins)	91.20%	10.91 ms
Proposed Custom 1D-CNN	DWT (db4, Level 4) + DEMON (128 bins)	94.93%	11.40 ms

The Capability of Feature Discrimination: Although traditional classifiers like SVM can create effective decision boundaries for classes that are clearly separated, they find it hard to differentiate between the acoustic objects that have overlapping spectra and similar shaft modulation frequency (the Fast Boat and the Rubber Zodiac, for instance). The innovative 1D-CNN addresses that disadvantage thanks to the hierarchical structure of convolution layers (with the receptive fields of 1×7 and 1×5) combined with Batch Normalization; thus this neural network can examine the nonlinear spatial relations between different frequency bins.

Trade-offs in Computational Latency: In terms of efficiency, while simple techniques like k-NN (10.05 ms) and SVM (10.91 ms) are more time-efficient considering their straightforward mathematical approaches, the deep 1D-CNN requires 11.40 ms for its forward pass.

Conclusion

The present study has shown that shallow water passive sonar target recognition can be done successfully by using the DWT method along with the DEMON algorithm and 1D-CNN to tackle the major issues related to high ambient noise and multipath problems. The findings of the experiment have suggested that the usage of the db4 wavelet for denoising has resulted in the effective enhancement of the signal by achieving very low RMSE value of 4.0195 while retaining the required transient acoustic modulations for the classification process, the custom 1D-CNN classifier model has achieved 94.93% accuracy and a macro-average F1-Score of 0.9505 which is better than conventional classification techniques, and the deep feature-based anomaly detection technique has been quite effective in identifying untrained acoustic signals, and their anomaly distances have exceeded the defined decision threshold.

Future work will focus on validating the proposed work using real-world empirical oceanic datasets and optimizing the neural network architecture for real-time, low-latency deployment on edge devices and Autonomous Underwater Vehicles (AUVs).

Author Contributions: Alnour: Conceptualization, study design, methodology, literature search, data extraction, quality assessment, formal data analysis, project administration, writing—original draft preparation, and writing—review and editing. **Abousetta:** Data curation, scientific supervision, critical review, writing—review and editing, and final manuscript approval. All authors have read and agreed to the published version of the manuscript and accept accountability for all aspects of the work.

Funding: This research received no external grant or financial funding.

Data Availability Statement: The synthetic passive sonar signal datasets and acoustic channel simulation models generated during this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgements: The authors express their sincere gratitude to the Libyan Academy for Graduate Studies for providing academic resources and technical support required to accomplish this research work.

References

[1] J. Benesty, J. Chen, and Y. Huang. *Microphone Array Signal Processing*. Berlin, Heidelberg: Springer-Verlag, 2008. https://doi.org/10.1007/978-3-540-78612-2

[2] B. Jensen, W. Kuperman, M. Porter, and H. Schmidt. *Computational Ocean Acoustics*, 2nd ed. New York, NY, USA: Springer, 2011. https://doi.org/10.1007/978-1-4419-8678-8

[3] W. Chung, A. Sutin, A. Sedunov, and M. Bruno. "DEMON acoustic ship signature measurements in an urban harbor." *Advances in Acoustics and Vibration*, vol. 2011, Art. no. 952798, 2011. https://doi.org/10.1155/2011/952798

[4] M. Wang, X. Wang, T. Fu, and M. Liu. "Passive tracking of underwater acoustic targets based on multi-beam LOFAR and deep learning." *PLOS ONE*, vol. 17, no. 12, p. e0273898, 2022. https://doi.org/10.1371/journal.pone.0273898

[5] I. Hummel, R. van der Mei, and S. Bhulai. "A survey on machine learning in ship radiated noise." *Ocean Engineering*, vol. 298, p. 117252, 2024. https://doi.org/10.1016/j.oceaneng.2024.117252

[6] G. Slamnoiu, C. Rujoiu, V. Iacob, and A. Cristea. "DEMON-type algorithms for determination of hydro-acoustic signatures of surface ships and of divers." *IOP Conference Series: Materials Science and Engineering*, vol. 145, no. 8, p. 082003, 2016. https://doi.org/10.1088/1757-899X/145/8/082003

[7] D. Darwish. "Improving techniques for convolutional neural networks performance." *European Journal of Electrical Engineering and Computer Science*, vol. 8, no. 1, pp. 1–16, 2024. https://doi.org/10.24018/ejece.2024.8.1.579

[8] O. Nacar, and T. Koc. "Enhancing CNN-Based Signal Denoising: A Novel Metric Framework with Harmonic Suppression through Hybrid Modeling." *IEEE Access*, vol. 13, 2025, In Press. https://doi.org/10.1109/ACCESS.2025.3534567

[9] T. Turov, A. Zhirnov, A. Pnev, and K. Stepanov. "Enhancing the distributed acoustic sensors' (DAS) performance by the simple noise reduction algorithms sequential application." *Algorithms*, vol. 16, no. 5, p. 217, 2023. https://doi.org/10.3390/a16050217

[10] R. Zaheer, M. Khan, and H. Kim. "A review on underwater beamforming: Techniques, challenges, and future directions." *Journal of Ocean Engineering and Science*, In Press, 2025. https://doi.org/10.1016/j.joes.2025.100123

[11] Y. Xie, J. Liu, and H. Wang. "DEMONet: Underwater acoustic target recognition based on multi-expert network and cross-temporal variational autoencoder." *arXiv preprint arXiv*, 2411.02758, 2024. https://arxiv.org/abs/2411.02758

[12] [P. Qian, J. Li, and S. Zhang. "Multi-Task Mixture-of-Experts Model for Underwater Target Localization and Recognition." *Remote Sensing*, vol. 17, no. 17, p. 2961, 2025. https://doi.org/10.3390/rs17172961

[13] L. Ogden, J. Perkins, and B. H. Houston. "Extraction of small boat harmonic signatures from passive sonar." *The Journal of the Acoustical Society of America*, vol. 129, no. 6, pp. 3768–3776, 2011. https://doi.org/10.1121/1.3578453

[14] Y. Yao, C. Sun, and X. Liu. "Application of waveguide invariant theory to analysis of interference phenomenon in deep ocean." *Acoustics*, vol. 2, no. 3, pp. 491–506, 2020. https://doi.org/10.3390/acoustics2030026

[15] V. Weinberg and M. Merwe. "Passive Sonar Sensor Placement for Undersea Surveillance." *arXiv preprint arXiv:2503.03940*, 2025. https://arxiv.org/abs/2503.03940

[16] S. Doan, T. Huynh-The, and D. Kim. "Underwater acoustic target classification based on dense convolutional neural network." *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2020. https://doi.org/10.1109/LGRS.2020.3029584

[17] A. Khan, M. Ali, and S. Lee. "Underwater target detection using deep learning: methodologies, challenges, applications and future evolution." *IEEE Access*, vol. 12, 2024. https://doi.org/10.1109/ACCESS.2024.3353688

[18] S. Jamal, J. Lakziz, Y. Benremdane, and S. Ouaskit. "Passive sonar detection and classification based on DEMON-LOFAR analysis and neural network algorithms." *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 15, no. 1, 2024. https://doi.org/10.5121/ijaia.2024.15106

[19] Y. Xincun, L. Ming, and Z. Tao. "Kalman filter applied in underwater integrated navigation system." *Geodesy and Geodynamics*, vol. 4, no. 1, pp. 46–50, 2013. https://doi.org/10.3724/SP.J.1246.2013.01046

[20] J. Jiang, X. Li, and Y. Chen. "Detection of underwater acoustic target using beamforming and neural network in shallow water." *Applied Acoustics*, vol. 189, p. 108626, 2022. https://doi.org/10.1016/j.apacoust.2021.108626

[21] P. Ashok and B. Latha. "Feature extraction of underwater acoustic signal target using machine learning technique." *Traitement du Signal*, vol. 41, no. 3, p. 1303, 2024. https://doi.org/10.18280/ts.410323

[22] X. Luo, Y. Zhang, and Z. Wu. "A survey of underwater acoustic target recognition methods based on machine learning." *Journal of Marine Science and Engineering*, vol. 11, no. 2, p. 384, 2023. https://doi.org/10.3390/jmse11020384

[23] X. Xiao, W. Xu, and Z. Gong. "Underwater acoustic target recognition using attention-based deep neural network." *JASA Express Letters*, vol. 1, no. 10, p. 106001, 2021. https://doi.org/10.1121/10.0006764

[24] D. Donoho. "De-noising by soft-thresholding." *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 1995. https://doi.org/10.1109/18.382009

[25] A. Cunha, M. Figueiredo, and C. J. Silvestre. "Simultaneous compression and denoising of side scan sonar images using the discrete wavelet transform." *OCEANS 2000 MTS/IEEE Conference and Exhibition*, vol. 1, pp. 403–408, 2000. https://doi.org/10.1109/OCEANS.2000.881268

[26] Z. Blal, R. Ali, and S. Yasser. "Improving and classification ECG signal using CNN by comparison signal processing techniques." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 2, no. 2, pp. 99–103, 2024. https://waujpas.com/index.php/journal/article/view/88

[27] F. Mishri and R. Ali. "Detection and Classification of Brain Tumours from MRI Images Using VGG16 Deep Learning Algorithm." *Wadi Alshatti University Journal of Pure and Applied Sciences*, vol. 3, no. 1, pp. 43–50, 2025. https://waujpas.com/index.php/journal/article/view/107