DOI: https://doi.org/10.62985/j.huit_ojs.vol26.no2E.406

A REVIEW OF UNDERWATER ACOUSTIC TARGET RECOGNITION IN THE ERA OF ARTIFICIAL INTELLIGENCE

Van Vuong Vu[1,*], Chi Hieu Ta[1], Truong Giang Bui[2], Ngoc Dong Nguyen[1]

[1]*Le Quy Don Technical University, Hanoi, Vietnam*

[2]*Naval Academy, Khanh Hoa, Vietnam*

*Email: *vuongvv@lqdtu.edu.vn*

Received: 12 November 2025; Revised: 8 March 2026; Accepted: 18 April 2026

ABSTRACT

Underwater Acoustic Target Recognition (UATR) is a critical technology for maritime domain awareness, naval defense, and oceanographic monitoring. With the increasing shift of naval operations toward complex littoral environments, conventional passive sonar systems that rely heavily on human operators and classical signal processing methods are encountering substantial difficulties. These challenges stem primarily from low signal-to-noise ratios (SNR), multipath propagation effects, and highly non-stationary background interference.

This paper presents a comprehensive review of the evolution of UATR technologies, progressing from traditional physics-based approaches such as LOFAR and DEMON to advanced deep learning frameworks. We critically analyze the limitations of conventional machine learning techniques (e.g., SVM and HMM) and evaluate how modern neural network architectures - including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), attention mechanisms, Transformers, and Generative Adversarial Networks (GANs) - have significantly enhanced feature extraction and classification performance in underwater environments.

Furthermore, this review addresses key contemporary challenges, including data scarcity, model interpretability, adversarial vulnerability, and the deployment of efficient models on resource-constrained edge devices such as sonobuoys. By synthesizing recent advancements, we highlight existing research gaps and propose future directions to develop more robust, intelligent, and practical underwater acoustic surveillance systems.

Keywords: Passive sonar, deep learning, feature extraction, adversarial robustness, edge computing.

1. INTRODUCTION

Passive sonar technology has a long history and continues to serve both military and civilian applications. It is widely used in oceanographic research, environmental monitoring, marine mammal tracking, and search-and-rescue operations. In the context of maritime surveillance and national security, passive sonar plays a vital role by detecting acoustic signals emitted from various underwater and surface sources, including submarines, surface vessels, marine animals, and natural phenomena.

These radiated signals are typically weak and heavily masked by ambient noise. They are further distorted by the complex underwater acoustic channel due to multipath propagation, frequency-dependent absorption, and varying sound speed profiles. Passive sonar systems capture these signals using hydrophone arrays. The raw acoustic data undergoes amplification and filtering to suppress unwanted noise, followed by analog-to-digital conversion. Subsequent processing employs classical digital signal processing techniques such as the Fourier Transform, Short-Time Fourier Transform (STFT), LOFAR

(Low Frequency Analysis and Recording), and DEMON (Detection of Envelope Modulation on Noise). These methods extract spectral and modulation features that enable target detection and classification based on acoustic signatures.

A major advantage of passive sonar is its stealth capability, as it does not emit any active signals, making it difficult for adversaries to detect the listening platform. However, it also has notable limitations, including the inability to detect silent targets and challenges in estimating target range when using a single hydrophone [1].

In recent years, the rapid advancement of artificial intelligence has brought transformative breakthroughs to underwater signal processing. Deep learning models can automatically learn discriminative features directly from raw signals or time-frequency representations, significantly reducing reliance on manual feature engineering. These AI-driven approaches not only improve classification accuracy but also enhance system adaptability across diverse and dynamic underwater environments. As a result, the integration of artificial intelligence into underwater acoustic target recognition (UATR) has become a critical research direction.

Underwater acoustic signals originate from a wide variety of sources, including human activities (ship propulsion, fishing operations) and natural phenomena (marine mammal vocalizations, seismic activity, wind, and waves). For instance, a moving vessel typically produces both broadband noise from propeller cavitation and hydrodynamic effects, as well as narrowband tonal components from machinery and propulsion systems.

Despite decades of research, underwater acoustic target classification remains highly challenging due to the complex ocean environment. Factors such as strong background noise, low signal-to-noise ratios (SNR), Doppler shifts, varying propagation conditions, and advancements in ship stealth technology continue to complicate reliable recognition. Traditionally, classification has depended heavily on the expertise of trained sonar operators using tools like LOFARgrams, DEMON spectra, beamforming, and power spectral density analysis. This human-centric approach demands constant attention and is susceptible to fatigue, environmental variations, and operator subjectivity.

While classical signal processing methods have provided a solid foundation, they often struggle to handle the non-stationary and highly cluttered nature of real-world underwater signals. In contrast, artificial intelligence-based techniques have demonstrated superior performance in extracting meaningful patterns from noisy data, making them particularly promising for robust target recognition in challenging maritime conditions.

This paper offers a comprehensive review of AI-based Underwater Acoustic Target Recognition. We systematically examine key aspects including feature extraction techniques and state-of-the-art classification algorithms. In addition, we analyze major challenges facing current systems and discuss promising directions for future research.

2. FEATURE EXTRACTION

UATR systems are generally built upon two primary stages: feature extraction and classification. Given the highly complex and dynamic nature of the underwater acoustic environment, extracting robust and discriminative features is essential for achieving reliable ship classification performance.

Over the years, researchers have developed a wide range of feature representations for UATR. These features can be broadly categorized into four main groups: time-domain, frequency-domain, time-frequency, and auditory-inspired features.

Time-domain features: Once digitized, underwater acoustic signals appear as one-dimensional time-series data representing amplitude changes over time. Raw waveform representations retain the maximum amount of original information. Common time-domain descriptors include zero-crossing rate and waveform statistics. However, these features are

often difficult to interpret physically and provide limited insight into the mechanical characteristics of ship sources. Consequently, they are rarely used in isolation for underwater target classification.

Frequency-domain features: To enhance physical interpretability and align with traditional sonar operator practices, frequency-domain features are commonly extracted using the Fourier Transform. These representations highlight the distribution of signal energy across different frequencies and reveal characteristic harmonic lines produced by ship machinery and propulsion systems. A key limitation, however, is their inability to capture temporal variations, which restricts their effectiveness in analyzing dynamic or transient acoustic events.

Time-frequency features: To address the shortcomings of purely time- or frequency-domain approaches, time-frequency representations have become the dominant choice in modern UATR. The most widely used method is the Short-Time Fourier Transform (STFT), which divides the signal into overlapping short windows to generate spectrograms. From these spectrograms, several specialized features are derived, such as cepstral coefficients, LOFAR spectra for narrowband low-frequency analysis [2], and DEMON spectra for extracting propeller modulation frequencies [3]. In addition, more advanced adaptive techniques including the Wavelet Transform (WT) and Hilbert–Huang Transform (HHT) have been employed to better handle non-stationary and nonlinear signals, especially those containing transient components or rapidly varying frequencies [4].

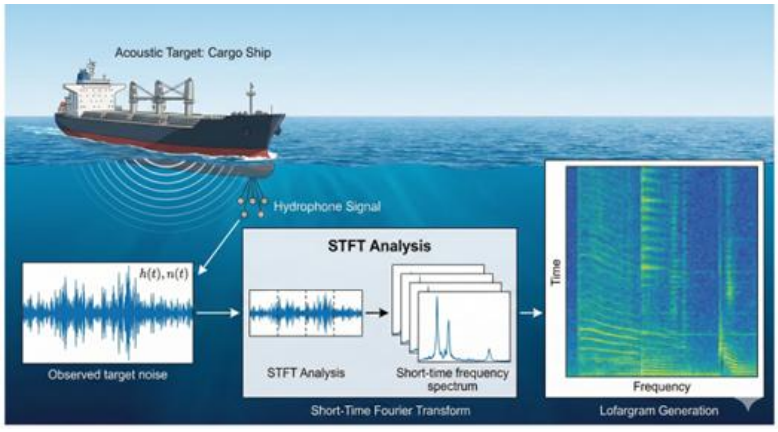

Fig 1. Workflow for Underwater Signal Processing and Lofargram Generation

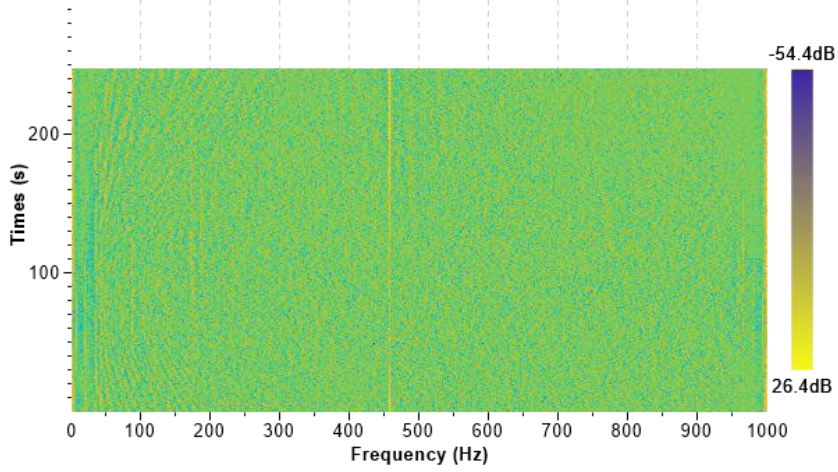

Fig 2. LOFARgram of the cargo ship SAMOS WARRIOR

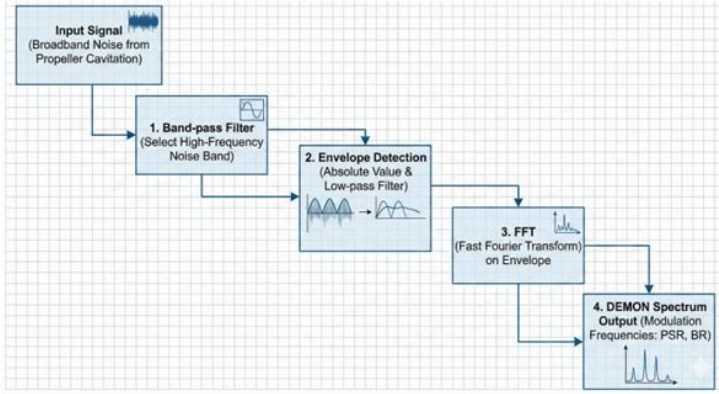

Fig 3. DEMON Processing Block Diagram

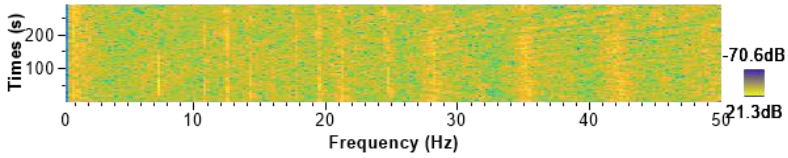

Fig 4. DEMON spectrum of the cargo ship SAMOS WARRIOR

Auditory-inspired features: Drawing inspiration from human auditory perception and the expertise of experienced sonar operators in distinguishing ship types, researchers have developed auditory-inspired features for UATR. Since human hearing processes frequency on a nonlinear scale, logarithmic frequency scales - particularly the Mel scale - are widely used to generate Log-Mel spectrograms. From these, Mel-Frequency Cepstral Coefficients (MFCCs) are extracted to create compact yet highly informative feature vectors that capture the most salient characteristics of the acoustic signal [5].

Similarly, Gammatone-based features, such as Gammatone Spectrograms (GST) and Gammatone-Frequency Cepstral Coefficients (GFCCs), have been proposed to more closely emulate the human cochlea's filtering mechanism. These features benefit from smoother filters and greater spectral overlap compared to traditional approaches. Additionally, Constant-Q Transform offers another effective logarithmic time-frequency representation, providing relatively consistent frequency resolution across different octaves.

In summary, each category of feature representations offers distinct advantages and limitations. The choice of the most suitable feature set depends heavily on the assumed stationarity of ship-radiated noise and the specific operational requirements of the UATR system.

3. UATR METHODS BASED ON MACHINE LEARNING

Machine learning (ML) techniques have been extensively employed in underwater acoustic target recognition owing to their relative simplicity, computational efficiency, and practicality for real-world deployment. A conventional ML pipeline typically consists of two steps: extracting handcrafted features from raw acoustic signals and feeding these features into shallow classifiers for target identification. These models learn decision boundaries by measuring similarity among features in high-dimensional space.

Among various ML algorithms, distance-based methods such as Support Vector Machines (SVMs) and k-Nearest Neighbors (KNN) have demonstrated strong performance in UATR tasks. SVMs, in particular, remain one of the most popular choices. By utilizing

kernel functions to project features into higher-dimensional spaces, SVMs can effectively separate different target classes with maximum margins. Several studies report that combining SVMs with carefully selected individual features yields good accuracy, while feature fusion strategies have been shown to further enhance performance on realistic underwater datasets [6]. KNN is valued for its straightforward implementation and competitive results, although it incurs high computational overhead due to pairwise distance calculations across the entire training set.

Probabilistic models, including Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs), offer another effective direction. HMMs are well-suited for modeling temporal sequences, whereas GMMs excel at representing static spectral distributions. However, these models can be difficult to train and are often sensitive to hyperparameter settings. Tree-based ensembles such as decision trees and Random Forests provide good interpretability and robustness against overfitting through bagging and feature randomness. Additionally, regression-based methods have been explored, especially in scenarios with limited labeled data.

In summary, traditional machine learning methods have played a significant role in the development of intelligent UATR systems and continue to serve as a solid foundation for more advanced approaches. Nevertheless, their relatively shallow representational capacity limits their ability to capture the intricate, non-linear patterns inherent in complex underwater acoustic environments [7], [8], [9].

4. UATR METHODS BASED ON DEEP LEARNING

The emergence of deep learning (DL) has fundamentally transformed Underwater Acoustic Target Recognition (UATR) by enabling end-to-end learning frameworks. In these models, feature extraction and classification are jointly optimized, eliminating the need for manual feature engineering. This section reviews prominent deep learning architectures that have been recently applied in the field.

4.1. Convolutional Neural Networks

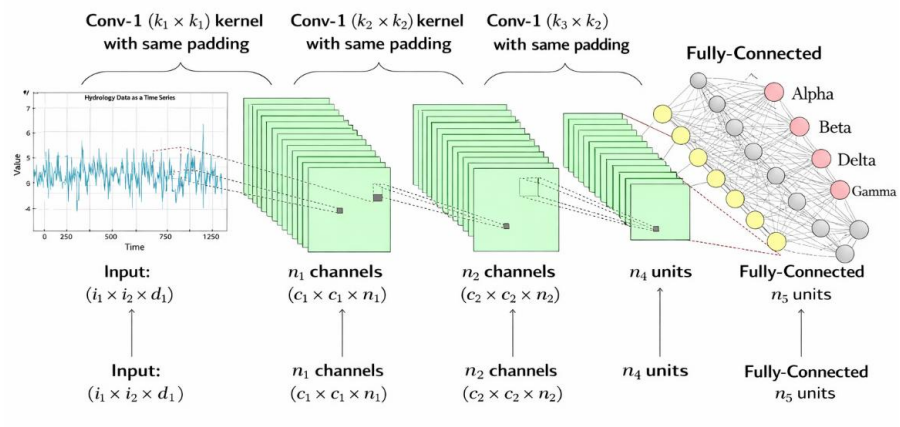

Fig 5. Typical CNN Architecture

Convolutional Neural Networks have gained widespread popularity in UATR due to their powerful ability to automatically learn hierarchical and discriminative features from complex acoustic data. In passive sonar applications, CNNs substantially reduce dependence on handcrafted features that traditionally require extensive domain expertise.

Time-frequency representations such as spectrograms, Log-Mel spectrograms, or Mel-Frequency Cepstral Coefficients (MFCCs) are commonly converted into two-dimensional images and fed as input to CNNs. Through successive convolutional layers, these networks effectively capture local patterns in both time and frequency domains, including narrowband spectral lines and modulation structures characteristic of different targets. Weight sharing and pooling operations help reduce the number of parameters, enhance noise robustness, and prevent overfitting even with limited training data.

Deeper CNN architectures, particularly those incorporating residual connections (e.g., ResNet), facilitate effective training by mitigating the vanishing gradient problem. CNN-based models have demonstrated excellent performance under low signal-to-noise ratio conditions typical of underwater environments. Compared to traditional classifiers such as SVMs or GMMs, CNNs generally deliver superior classification accuracy. Despite the persistent challenge of limited labeled data, CNNs remain one of the most effective and foundational architectures for underwater acoustic target recognition [7], [9].

4.2. Recurrent Neural Networks

Recurrent Neural Networks (RNNs) represent a powerful class of deep learning models specifically designed to handle sequential and time-varying data. In the field of underwater acoustic target recognition, RNNs are particularly well-suited because they can effectively model temporal dependencies present in sonar signals.

Unlike Convolutional Neural Networks (CNNs), which primarily capture local spatial patterns, RNNs maintain a memory of previous time steps. This enables them to learn dynamic characteristics such as changes in frequency content or amplitude over time features that are critical for distinguishing ship-radiated noise. However, vanilla RNNs frequently encounter training difficulties due to vanishing or exploding gradients, especially when processing long sequences.

To overcome these issues, advanced variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been widely adopted. These architectures incorporate gating mechanisms that selectively retain or forget information, allowing the network to capture long-term dependencies commonly found in underwater acoustic signals. In practice, RNNs are often applied to sequences of features derived from time-frequency representations like spectrograms or MFCCs. Hybrid models that combine CNNs and RNNs benefit from both spatial feature extraction and temporal modeling, leading to improved overall recognition performance.

Although RNN-based approaches generally require higher computational resources than standalone CNNs, they remain essential for effectively capturing the temporal dynamics of underwater acoustic signals [7], [9].

4.3. Attention Networks

Attention mechanisms have gained significant traction in underwater acoustic target recognition systems. They enhance model performance by enabling the network to dynamically focus on the most relevant parts of the acoustic signal while reducing the impact of background noise.

In the challenging underwater environment, where signals are often weak and heavily corrupted by noise, attention modules assign higher importance weights to discriminative time-frequency regions such as characteristic spectral lines and propeller modulation patterns that are strongly associated with specific vessel types. When integrated with common inputs like spectrograms or MFCCs, attention layers effectively highlight target-specific features and suppress irrelevant noise components.

Numerous studies have shown that incorporating attention mechanisms into CNN or hybrid CNN-RNN architecture leads to notable improvements in classification accuracy and generalization ability, especially in low signal-to-noise ratio conditions and when training data is limited. By modeling global dependencies across features, attention-based approaches have become a key component in state-of-the-art UATR systems [7], [9].

4.4. Transformer

Transformer architecture, which relies primarily on self-attention mechanisms, has emerged as a powerful deep learning model for capturing global dependencies in sequential data. In underwater acoustic target recognition, Transformers are especially well-suited to handle the complex and non-stationary characteristics of sonar signals.

When applied to time-frequency representations such as spectrograms, Log-Mel spectrograms, or MFCC sequences, Transformers can effectively model long-range relationships across distant time frames and frequency bands without the sequential processing constraints of traditional RNNs. This capability allows the model to emphasize the most informative frequency components and temporal patterns that are highly discriminative for different target types.

Recent research indicates that Transformer-based models and hybrid CNN-Transformer architecture often achieve competitive or even superior performance compared to conventional CNNs in UATR tasks. However, they typically demand large amounts of training data and incur substantial computational costs due to the quadratic complexity of the self-attention mechanism. Despite these limitations, Transformers are considered a highly promising direction for the development of next-generation underwater acoustic target recognition systems [7], [9].

5. CHALLENGES AND PROPOSED SOLUTIONS

Intelligent underwater acoustic target recognition systems face several critical challenges that limit their real-world effectiveness.

One of the most pressing issues stems from the harsh underwater environment itself. Strong background noise, complex acoustic propagation paths, and severe multipath effects result in very low signal-to-noise ratios (SNRs), making it extremely difficult to separate weak target signals from ambient interference. Moreover, collecting high-quality labeled underwater acoustic data is expensive and technically demanding, leading to significant data scarcity especially in few-shot learning scenarios [3]. These constraints seriously impair the model's ability to learn robust and discriminative features.

To tackle these problems, researchers have proposed several promising approaches. These include joint training frameworks that integrate denoising and recognition through cross-attention fusion, as well as biologically inspired loss functions that emulate human auditory attention. Advanced preprocessing techniques such as dual-path noise reduction and adaptive filtering are also used to enhance input signal quality. For data-limited situations, techniques like Siamese networks [3], self-supervised learning (contrastive and generative methods), and meta-learning strategies have shown effectiveness in improving feature learning under challenging conditions.

Beyond environmental difficulties, model interpretability remains a major concern. Deep learning models are often regarded as "black boxes" [10], making it hard to understand which features drive their decisions and whether temporal or frequency-domain patterns dominate the classification process. This challenge is further intensified by the non-stationary, non-Gaussian, and nonlinear nature of underwater acoustic signals. Current efforts to improve interpretability mainly rely on visualization tools such as Grad-CAM,

dimensionality reduction techniques like t-SNE, and attention maps. However, these methods are largely borrowed from computer vision and may not fully capture the unique physical characteristics of acoustic data. As a result, graph-based modeling - representing signals or features as graphs to reveal complex correlations - has emerged as a promising direction for more physically meaningful interpretability.

Another important limitation is the poor generalization of UATR systems. In practical deployments, models frequently encounter new sea areas, different noise profiles, and previously unseen vessel types, causing significant performance drops due to domain shifts (environmental mismatch). To address this, data augmentation, transfer learning, and pre-training on large cross-domain datasets (e.g., ImageNet or AudioSet) are commonly employed [11]. Efficient adaptation techniques such as LoRA, QLoRA, and Adapter Tuning allow models to quickly adjust to new environments. Additionally, multi-scale feature fusion and the integration of raw waveforms with frequency-domain representations further improve robustness.

Finally, adversarial robustness presents a serious security concern, particularly for military and safety-critical applications. Deep learning models are vulnerable to carefully designed adversarial perturbations that can cause severe misclassifications (e.g., mistaking a small boat for a large tanker) with imperceptible changes to the input. Strengthening UATR systems against such attacks through adversarial training, adversarial example detection, and defensive distillation is therefore essential. Evaluating model robustness under strong adversarial conditions should become standard practice before real-world deployment [12].

6. CONCLUSION

This review has provided a systematic overview of Underwater Acoustic Target Recognition amid the rapid progress of artificial intelligence. We have traced the development trajectory from classical physics-based signal processing techniques to contemporary deep learning approaches, highlighting both the strengths and shortcomings of current methodologies.

Traditional feature extraction methods such as LOFAR and DEMON, together with conventional machine learning classifiers, have laid a solid foundation for passive sonar systems. However, they continue to face significant difficulties in dealing with low signal-to-noise ratios, non-stationary signals, and the complex propagation characteristics of the underwater environment.

The advent of deep learning has markedly advanced the field by introducing end-to-end frameworks capable of automatically learning hierarchical and highly discriminative features directly from raw signals or time-frequency representations. Modern architectures - including CNNs, RNNs, attention mechanisms, and Transformers - have achieved superior performance in processing noisy and dynamic underwater acoustic signals. Recent research trends focusing on data scarcity [13], domain generalization, model interpretability, adversarial robustness, and efficient edge deployment reflect a clear shift toward more practical and reliable intelligent surveillance systems.

Nevertheless, several critical challenges persist. These include the scarcity of labeled underwater acoustic datasets, environmental mismatches across different sea regions, the inherent black-box nature of deep learning models, and susceptibility to adversarial attacks. These issues continue to constrain widespread real-world implementation.

Future research efforts should prioritize the integration of domain-specific underwater acoustic knowledge with data-driven models. Promising directions include the development of self-supervised and few-shot learning frameworks, the creation of physically grounded

interpretable models, and the design of lightweight yet robust architectures suitable for real-time deployment on resource-limited platforms such as sonobuoys and autonomous underwater vehicles.

By successfully addressing these challenges, intelligent UATR systems are poised to play an increasingly important role in maritime security, ocean monitoring, and autonomous naval operations in the era of artificial intelligence.

REFERENCES

[1] B. T. Giang, *Underwater Acoustics Theory and Applications*, Nha Trang: Naval Academy Publishing House, 2021.

[2] G. Jin, F. Liu, H. Wu, and Q. Song, "Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 32, no. 2, pp. 205–218, 2020, doi: 10.1080/0952813X.2019.1647560.

[3] D. Liu, W. Shen, W. Cao, W. Hou, and B. Wang, "Design of Siamese network for underwater target recognition with small sample size," *Applied Sciences*, vol. 12, no. 20, Art. no. 10659, 2022, doi: 10.3390/app122010659.

[4] X. Zeng and S. Wang, "Underwater sound classification based on Gammatone filter bank and Hilbert-Huang transform," in *Proc. IEEE Int. Conf. Signal Processing, Communications and Computing (ICSPCC)*, Guilin, China, 2014, pp. 707–710, doi: https://doi.org/10.1109/ICSPCC.2014.6986287.

[5] G. Hu, K. Wang, and L. Liu, "Underwater acoustic target recognition based on depthwise separable convolution neural networks," *Sensors*, vol. 21, no. 4, Art. no. 1429, 2021, doi: https://doi.org/10.3390/s21041429.

[6] J. Liu, Y. He, Z. Liu, and Y. Xiong, "Underwater target recognition based on line spectrum and support vector machine," in *Proc. 2014 Int. Conf. Mechatronics, Control and Electronic Engineering (MCE)*, Shenyang, China, 2014, pp. 79–84, doi: https://doi.org/10.2991/mce-14.2014.17.

[7] S. Feng, S. Ma, X. Zhu, M. Yan, and H. Xu, "Artificial intelligence-based underwater acoustic targets recognition: A survey," *Remote Sensing*, vol. 16, no. 17, Art. no. 3333, 2024, doi: https://doi.org/10.3390/rs16173333.

[8] J. Choi, Y. Choo, and K. Lee, "Acoustic classification of surface and underwater vessels in the ocean using supervised machine learning," *Sensors*, vol. 19, no. 16, Art. no. 3492, 2019, doi: https://doi.org/10.3390/s19163492.

[9] N. Müller, J. Reermann, and T. Meisen, "Navigating the depths: A comprehensive survey of deep learning for passive underwater acoustic target recognition," *IEEE Access*, vol. 12, pp. 154092–154118, 2024, doi: https://doi.org/10.1109/ACCESS.2024.3480788.

[10] Y. Liang, S. Li, C. Yan, M. Li, and C. Jiang, "Explaining the black-box model: A survey of local interpretation methods for deep neural networks," *Neurocomputing*, vol. 419, pp. 168–182, 2021, doi: https://doi.org/10.1016/j.neucom.2020.08.011.

[11] D. Li, F. Liu, T. Shen, L. Chen, X. Yang, and D. Zhao, "Generalizable underwater acoustic target recognition using feature extraction module of neural network," *Applied Sciences*,

vol. 12, no. 21, Art. no. 10804, 2022, doi: https://doi.org/10.3390/app122110804.

[12] S. Feng, X. Zhu, S. Ma, and Q. Lan, "Adversarial attacks in underwater acoustic target recognition with deep learning models," *Remote Sensing*, vol. 15, no. 22, Art. no. 5386, 2023, doi: https://doi.org/10.3390/rs15225386.

[13] M. Irfan, Z. Jiangbin, S. Ali, M. Iqbal, Z. Masood, and U. Hamid, "DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification," *Expert Systems with Applications*, vol. 183, Art. no. 115270, 2021, doi: https://doi.org/10.1016/j.eswa.2021.115270.