frontiers | Frontiers in Physics

TYPE Mini Review
PUBLISHED 17 November 2025
DOI 10.3389/fphy.2025.1682253

OPEN ACCESS

EDITED BY
Taehwa Lee,
Toyota Motor North America, United States

REVIEWED BY
Chen Shen,
Rowan University, United States

*CORRESPONDENCE
Hongjian Jia,
jiahongjian@hlju.edu.cn

RECEIVED 08 August 2025
REVISED 17 September 2025
ACCEPTED 05 November 2025
PUBLISHED 17 November 2025

CITATION
Xu T, Jia H and Qin J (2025) Explainable underwater target recognition models: principles, methods, and applications. *Front. Phys.* 13:1682253. doi: 10.3389/fphy.2025.1682253

Explainable underwater target recognition models: principles, methods, and applications

Tianyang Xu[1], Hongjian Jia[1]* and Jixing Qin[2]

[1]Electrical Engineering College, Heilongjiang University, Harbin, China, [2]State Key Laboratory of Acoustics and Marine Information, Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

With the increasing strategic importance of the ocean, underwater intelligent systems have become essential for signal processing, target recognition, and autonomous navigation. The widespread application of deep learning has significantly advanced underwater acoustic missions, but its "black box" nature has led to critical concerns about decision explainability, limiting its trustworthy application in high-risk scenarios. This paper provides a systematic review of explainable models for underwater target recognition, elaborating on the core concepts and main methods of explainability. It also reviews research progress and representative achievements in sonar imaging, signal analysis, and autonomous navigation. Finally, future directions, including causal reasoning, cross-modal collaboration, and physical knowledge integration, are identified to provide a reference for developing safe and reliable underwater intelligent systems.

KEYWORDS

underwater intelligent perception, underwater target recognition, artificial intelligence, explainable artificial intelligence, explainability in deep learning

1 Introduction

With the accelerating development of marine resources, the demand for underwater intelligent perception continues to grow. Driven by deep learning, underwater target recognition has made significant progress in efficiency, accuracy, and automation. However, its performance relies on large-scale neural networks and suffers from the "black box" problem, which limits the model applications in real-world scenarios. Especially in complex underwater environments, unstable propagation paths, strong noise, and diverse target shapes make the reliability of model outputs directly influence combat decision-making, autonomous navigation, and anomaly response. Therefore, the introduction of explainable mechanisms has become a key way to improve system stability, enhance human-machine collaboration, and cope with environmental uncertainty. Explainability not only enhances the interpretability of the model but also provides strong support for performance optimization, algorithm review, and feature visualization.

To address these challenges, Longo et al. [1] proposed a research roadmap for XAI (Explainable Artificial Intelligence) through which nine major categories and 27 questions derive pertinent scientific inquiries under deep reflection for the current positioning about explainability. This paper begins with the definition of explainability, reviews mainstream modeling approaches and representative

application cases, and focuses on recent advances in explainable modeling for underwater tasks. Finally, this paper summarizes the current bottlenecks regarding underwater modeling and highlights the most promising future development paths involving causal reasoning, cross-modality collaboration, and physical knowledge integration.

2 Fundamental concepts and motivations for explainability

The trustworthiness of artificial intelligence results hinges not only on whether they "can be done correctly," but also on whether they can "explain why they are correct." The concept of explainability is formalized in this context and systematically developed, focusing primarily on the reasoning behind the decisions or predictions made by artificial intelligence algorithms.

Doshi-Velez and Kim [2] state that "explainability is a functional property of a model such that humans can understand its behavior." Explainability is explaining something in comprehensible terms. Therefore, in machine learning systems, explainability is defined as the ability to explain or present in understandable terms to humans. Explainability can be evaluated in two main ways: one through contextual assessment within specific applications, and the other through quantifiable proxy metrics. In supervised learning, Lipton [3] refines the properties of explainable models into two categories. The first category is model transparency, the opposing attribute of "black-box" systems, encompassing model-level explainability such as simulatability [4], parameter decomposability at the single-component level [5], and training algorithm transparency [6]. The second category is *post hoc* explainability, which involves analyzing, reconstructing, or visualizing black-box models through additional methods. Explainability is crucial for building trust in artificial intelligence, assisting with audits and diagnostics, enhancing human-machine collaboration, and addressing regulatory requirements. Existing research still lacks the ability to generate reliable explanations in real time, achieve two-way interaction, and establish interpretability assessment standards. Further research is needed to bridge the gap between theory and application.

In the practical underwater problems, the explainable demand is particularly prominent. On the one hand, the operating environment is complex and changeable, and the channel interference is serious, so the correctness of the system decision-making results will have a great impact on combat command, autonomous navigation and other tasks; On the other hand, limited by the observation mode and communication bandwidth, the cost of manual intervention is extremely high, and the system urgently needs the ability of "self-explanation and self-diagnosis" to support its stable operation in an uncontrollable environment. In this context, it has become a key research topic to develop an explainable model suitable for underwater perception and decision-making tasks. As shown in Figure 1, it can be seen that AUV (Autonomous Underwater Vehicle) platforms and control centers rely on sonar and artificial intelligence in sensing and decision-making. The opacity of AI models has invoked concern and thus triggered research into the reasoning of complex AI models to improve their transparency, reliability, and controllability in carrying out complex ocean missions.

3 Categorization of explainable modeling approaches

To address explainability challenges posed by ensemble learning and deep neural networks, Arrieta et al. [7] systematically reviewed core concepts, classifications, and methods of explainable models. They proposed a user-centered definition of explainability and outlined both transparent models and *post hoc* explanation techniques for black-box systems. Deck et al. [8] highlighted the limitations of current fairness claims in XAI and argued that explainability should be seen as one of many socio-technical tools, not a cure-all. Giannini et al. [9] introduced a unified mathematical framework from a category theory perspective to formalize diverse explanation paradigms and strengthen the theoretical foundations of XAI.

Building upon that, explainable models can generally be split into two broad subgroups based on the criterion of architecture integration: intrinsically interpretable models and *post hoc* explainability methods. Intrinsically interpretable models are those in which design intent goes toward achieving their explainability feature through their simplicity, transparent inference path, and clarity of feature attribution. Classic examples are logistic regression, decision trees, rule-based learning, and naive bayes classifiers.

Nevertheless, intrinsically interpretable models struggle with high-dimensional nonlinear data, such as underwater acoustic signals, which suffer interference from reverberation and multipath. To address the above challenges, a recent trend incorporates a white-box-black-box approach in one cohesive framework. This relies on the input-output behavior of the already trained black-box models to trace the feature contribution responsible for a prediction or explain the region of interest. When machine learning models lack transparency, additional methods are employed to explain their decision processes, constituting the essence of *post hoc* explainability techniques [7]. Examples of such *post hoc* explainability methods include natural language explanations [10], model visualization [11], local interpretable techniques [12], and instance-based explanations [13]. Specific examples of these methods are gradient-weighted class activation mapping (Grad-CAM) [14], shapley additive explanations (SHAP) [15], and local interpretable model-agnostic explanation (lime) [16].

These methods are presented in the form of image heat maps, feature rankings or explanation maps. They not only maintain the accuracy requirements of target detection, but also meet the auditability requirements of the task through rule extraction, providing a systematic solution to enhance model explainability and facilitate practical deployment.

4 Research on explainable models in underwater target recognition

Underwater target recognition is a critical technology in fields such as marine exploration, underwater navigation, and military defense. Due to the highly complex underwater environment,

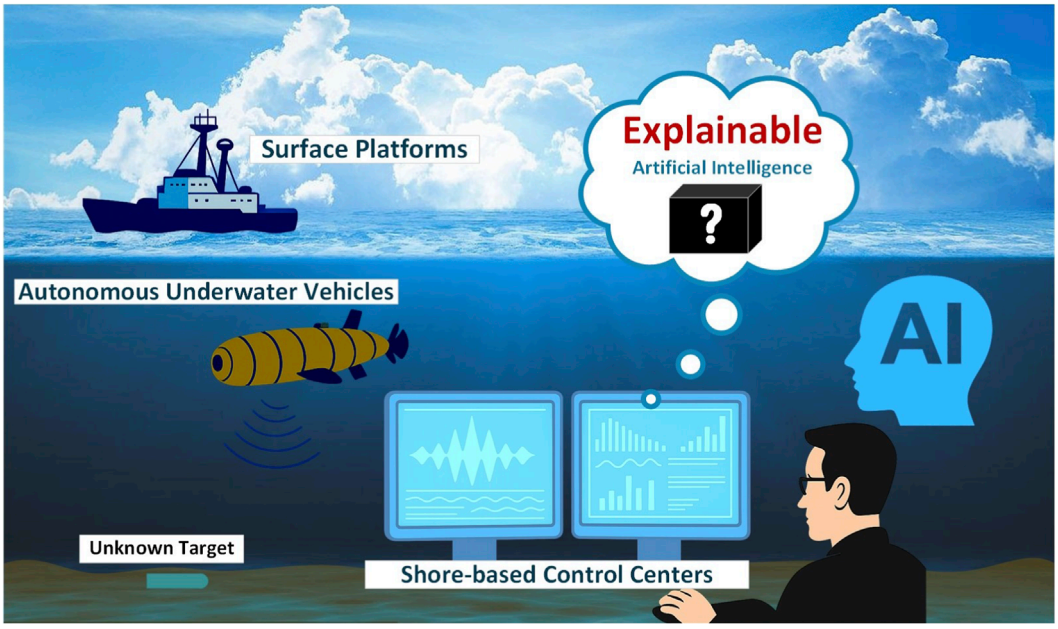

FIGURE 1
Overview of explainability challenges and core tasks in underwater intelligent systems.

diverse target shapes and materials, and the scarcity of effective training data (with high acquisition costs), many challenges remain to be addressed. With the development of stealth materials and advanced manufacturing techniques, underwater targets are becoming quieter and more varied, posing new difficulties for traditional recognition models that rely on single features. Research on explainable models for underwater targets not only helps improve recognition accuracy but also provides cognitive support for decision-making in complex scenarios, making it a key research direction in recent years. This section reviews explainable techniques for underwater tasks from four representative perspectives.

4.1 Sonar image classification and recognition

In the past, the detection and classification of early sonar images required manual participation, and the detection results were not real-time. With the advent of deep learning, several technologies, such as CNN [17] and transfer learning [18], improved the recognition performance of side-scan sonar systems. However, the transfer processes mixed abstract features shared among the source task and the target task, thus complicating the decision-making task. The "black box" nature of the model limits its credibility in high-security scenarios such as military defense. Hence, explainability becomes a crucial demand.

Jin et al. [19] introduced *post hoc* explainability methods at the preprocessing stage of Echoscope sonar data, applying visual attention to target area extraction. Through ImageNet pre-training and transfer learning strategy, they achieved a recognition accuracy of 97.3% for 9 categories of underwater targets. Cheng et al. [20] integrated global attention mechanisms into the YOLO network to tackle the problem of small targets being detected in side-scan sonar images. Walker et al. [21] presented a method to promote sonar SAS image classification explainability via a systematic analysis framework that involved both quantitative (divergence metrics) and qualitative (lime) approaches. Richard et al. [22] compared the consistency between XAI explanations and sonar operators' cognition and pointed out the future research direction of XAI that combines fuzzy logic or semantics. In the case of mine hunting, *post hoc* explainability methods were applied to provide explanations for the decisions made by complex neural networks. Keenan et al. [23] put forward a low-complexity human-machine collaborative classification framework using transfer learning, wherein RISE was proved capable of quantitatively measuring model improvements, along with the associated SR metric, offering a new tool for the AI trustworthiness study of sonar image analysis. Xie et al. [24] proposed a sonar image classification framework based on a feature-fusion attention network, using dual attention mechanisms (channel + spatial) to focus on key feature regions. It used Grad-CAM heatmaps to visualize target contours significantly, improving accuracy over traditional methods and offering innovative insight into sonar detection result explainability.

4.2 Explainable analysis of underwater target signal characteristics

Feature extraction is a critical component of underwater target detection and recognition. However, the complexity of underwater acoustic signal propagation often limits the effectiveness of traditional approaches. In recent years, explainability techniques have increasingly been integrated into deep learning models,

enhancing the transparency and reliability of tasks such as feature extraction, target tracking, and classification.

Zhu et al. [25] deeply coupled the physical mechanisms of ship-radiated noise with sub-band signal features. Using Grad-CAM heatmaps, they showed that low-frequency networks are focused on line spectra due to propeller rotation noise, while high-frequency networks are focused on broadband continuous spectra due to cavitation noise. This created a new, explainable paradigm that combines domain knowledge with deep learning. Wang et al. [26] used SHAP to provide local explanations of abnormal signals in ship data, then applied lime to identify the key features of abnormal signals most influential on the model, supporting user understanding of model decision-making and root cause analysis. Kubicek [27] integrated physics domain knowledge with feature engineering while applying Grad-CAM to explain CNN's decision-making basis to improve the transparency and trustworthiness of the classification model. The team associated features with resonance scattering theories (e.g., specular reflection and Rayleigh wave) and proposed a method of combining acoustic models to screen signal features with clear physical meaning, achieving accurate classification of elastic spherical shells of four different materials. Du et al. [28] applied Bayesian deep learning to classify active and passive sonar data. The Bayesian framework provides prediction confidence (such as entropy value) through posterior inference and supports decision interpretation, which helps the model perform more robust detection and improves the management of noise and uncertainty. Domingos et al. [29] pointed out that the underwater acoustic data classification model not only needs high accuracy, but also must have an explainable reasoning mechanism. Investigating the correlation between acoustic physical models and deep learning can enhance model explainability in complex environments. Feng et al. [30] noted that the model in the field of underwater acoustic target recognition is still in the initial stage, mainly relying on the *post hoc* explainability methods. Grad-CAM, t-SNE, and attention mechanism can reveal the attention area and feature distribution of the model to a certain extent, but it is difficult to reflect the acoustic and physical laws behind it. In the future, a new method combining a graph model with physical priors is still needed.

4.3 Explainable mechanisms for autonomous underwater vehicle decision-making

Autonomous surface and underwater vehicles (ASVs/AUVs) have become an indispensable tool in ocean missions. Its path planning and navigation strategies are usually generated by complex deep reinforcement learning models, but the model lacks transparency, which makes it difficult to understand its decision-making process.

Veitch et al. [31] proposed a human-centered XAI framework for Autonomous Surface Vehicles, emphasizing the distinction between human-centered and technology-centered XAI approaches and outlining design directions for human-computer interaction in autonomous systems. The framework includes three core cognitive processes: reducing user cognitive load through analogies, enhancing usability through visualization, and building user trust through mental simulation. Qiao et al. [32] identified insufficient explainability in ASV deep learning models and proposed feature visualization techniques to enhance understanding of CNN decision-making processes. They found a direction for model simplification by replacing parts of the deep network with shallow networks. Chen et al. [33] summarized the key challenges of the explainability requirement of unmanned surface vehicle (USV) motion control model, including how to establish the explainable relationship between simulation data and decision-making, how to explain how to reward function drives strategic behavior, and how to understand the logic of reinforcement learning decision-making.

Yan et al. [34] integrated binocular vision with an improved deep reinforcement learning algorithm. By analyzing the weights and design logic of various reward functions, they made clear the optimization goals of the model, which addressed motion planning for AUVs in uncertain model parameters and the complex underwater environment. Liu et al. [35] established an understandable geometric foundation in the UUV positioning algorithm, and transformed the forward-looking sonar positioning problem into an intuitive spatial geometric constraint problem, which made the behavior of the target positioning algorithm based on UUV forward-looking sonar clear and explainable. Aubard et al. [36] pointed out that explainable AI and model verification are the core of safe AI, and the limited sonar data aggravates the difficulty of model explainability. In the future, it is necessary to build a standardized sonar data platform to promote the autonomy and safety of AUVs in complex underwater environment.

4.4 Explainable fusion mechanisms for multimodal perception

In underwater target recognition, the system's recognition performance is limited under the condition of single-physics field detection. Multi-source information fusion model makes use of the complementarity of different modal information, such as multi-sensor, multi-platform and multi-detector to improve the system's ability to distinguish underwater targets. However, there is also a new black-box problem in the fusion of different modal data.

Sun et al. [37] combed the evolution of multi-modal explainable AI technology from three dimensions: data explainability, model explainability and post-processing explainability, and pointed out the development direction of visual data bottleneck breakthrough and truth-free environment interpretation, but the application in underwater scenes lacked systematic analysis. Li et al. [38] adopted a feature-level information fusion strategy, combined with multi-channel CNN-Transformer network to complete multi-scale feature extraction, and realized high-precision underwater target recognition under the condition of low signal-to-noise ratio. Cai et al. [39] built a detection system based on underwater acousto-optic multimodal cooperation, which transformed the camera results into auxiliary constraints for sonar detection, and adopted the improved Faster R-CNN framework to realize cooperative detection. Zheng et al. [40] combined the feature selection method based on the visualization technology (Grad-CAM) with the local feature enhancement method based on the sub-regional channel aggregation net (SRCA-Net). The study provided a visual basis for acoustic feature selection by fusing the network model. Zhang

TABLE 1 Overview of explainable underwater target recognition models by explainability methods.

XAI method	Category	Application	Contribution	Reference
Attention mechanism	Post-hoc analysis	Sonar image classification ASV human-computer interaction	DCNN + transfer learning Omni-dimensional dynamic convolution Human-centered models with analogy and visualization	[19, 20, 31]
Lime	Post-hoc analysis	Sediment classification; mine hunting; human-machine collaboration	Kullback-leibler divergence CNN; randomized input sampling	[21, 22, 23]
Grad-CAM	Post-hoc analysis + intrinsically interpretable model	Sonar image classification; ship-radiated noise recognition; Underwater target recognition	Dual attention mechanism + transfer learning; heatmap in different frequency bands; CNN + scattering mechanisms Feature enhancement + SRCA-net	[24, 25, 27, 40]
SHAP	Post-hoc analysis	Ship abnormal signal detection	LSTM + auto-encoder model	[26]
Bayesian deep learning	Intrinsically interpretable model	Underwater target detection	Generative models	[28]
Process explainability methods	Intrinsically interpretable model	USV motion control AUV motion planning UUV target positioning	Markov decision process + reinforcement learning; interruption-driven explainability Plane intersection + direction distance constraint	[33, 34, 35]
Feature visualization	Post-hoc analysis	Ship-radiated noise recognition; underwater target detection; AUV fault detection	Feature fusion + CNN-transformer network Acoustic and optical information fusion + faster R-CNN; multi-sensor feature fusion + KAN	[38, 39, 41]

et al. [41] used multi-sensor feature fusion KAN network (MFKAN) to realize the effective fusion of feature levels and enhance the expression ability of fault features in the background of marine noise. Nevertheless, the challenges of real-time and cross-modal consistency will still be faced in the future. Table 1 summarizes an overview of explainable underwater target recognition models organized by explainability methods.

5 Discussion

This paper reviews the concepts, origins, and methods of XAI, and explores the progress and challenges of its application in sonar recognition, signal analysis, AUV navigation, and multimodal fusion. Underwater explainable models not only provide explanations but also aim to ensure that systems are usable in complex missions, trustworthy under uncertain conditions, and controllable in high-risk scenarios. This is crucial for improving the safety and effectiveness of underwater missions and preparing for future operations in harsh environments. To construct a reliable and intelligent underwater recognition system, not only achieving high accuracy but also improving aspects of traceability, explainability, and control is mandatory. Recent studies involving explainable models for various underwater tasks have been leading toward more profound integration and collaboration on several fronts. Future research directions of interpretation include multimodal fusion (unification of acoustic, optical and magnetic data), causal drive (revealing variable mechanisms), human-computer collaboration (interactive on-demand explanations), and scenario adaptation (multi-granularity and multi-format presentation). All these efforts together will form the promising future of explainable and applicable underwater intelligent systems.

Author contributions

TX: Writing – original draft, Investigation, Resources. HJ: Funding acquisition, Writing – review and editing. JQ: Writing – review and editing, Supervision.

Funding

The authors declare that financial support was received for the research and/or publication of this article. This work was supported by the State Key Laboratory of Acoustics, Chinese Academy of Sciences (Grant No.SKLA202402), the Opening Fund of National Key Laboratory of Underwater Acoustic Technology (Grant No.SSKF202406) and the Fundamental Research Funds for

the Higher Institutions in Heilongjiang Province (Grant Nos.2024-KYYWF-0095 and 2024-KYYWF-0101).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Longo L, Brcic M, Cabitza F, Choi J, Confalonieri R, Ser JD, et al. Explainable Artificial intelligence (XAI) 2.0: a manifesto of open challenges and interdisciplinary research directions. *Inf Fusion* (2024) 106:102301. doi:10.1016/j.inffus.2024.102301

2. Doshi Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *stat* (2017) 1050:2. doi:10.48550/arXiv.1702.08608

3. Lipton ZC. The mythos of model interpretability: in machine learning, the concept of interpretability is both important and slippery. *Queue* (2018) 16(3):31–57. doi:10.1145/3236386.3241340

4. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" explaining the predictions of any classifier. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (2016). p. 1135–44. doi:10.1145/2939672.2939778

5. Lou Y, Caruana R, Gehrke J, Hooker G. Accurate intelligible models with pairwise interactions. In: *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (2013). p. 623–31. doi:10.1145/2487575.2487579

6. Mahendran A, Vedaldi A. Understanding deep image representations by inverting them. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015). p. 5188–96. doi:10.1109/CVPR.2015.7299155

7. Arrieta AB, Díaz Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* (2020) 58:82–115. doi:10.1016/j.inffus.2019.12.012

8. Deck L, Schoeffer J, De-Arteaga M, Kühl N. A critical survey on fairness benefits of explainable AI. In: *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency* (2024). p. 1579–95. doi:10.1145/3630106.3658990

9. Giannini F, Fioravanti S, Barbiero P, Tonda A, Liò P, Di Lavore E. Categorical foundation of explainable AI: a unifying theory. In: *World conference on explainable artificial intelligence*. Cham: Springer Nature Switzerland (2024). p. 185–206. doi:10.48550/arXiv.2304.14094

10. Krening S, Harrison B, Feigh KM, Isbell CL, Riedl M, Thomaz A. Learning from explanations using sentiment and advice in RL. *IEEE Trans Cogn Developmental Syst* (2016) 9(1):44–55. doi:10.1109/TCDS.2016.2628365

11. Dimitriadis G, Neto JP, Kampff AR. t-SNE visualization of large-scale neural recordings. *Neural Comput* (2018) 30(7):1750–74. doi:10.1162/neco_a_01097

12. Denton EL, Zaremba W, Bruna J, LeCun Y, Fergus R. Exploiting linear structure within convolutional networks for efficient evaluation. *Adv Neural Inf Process Syst* (2014) 27. doi:10.5555/2968826.2968968

13. Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J. Distributed representations of words and phrases and their compositionality. *Adv Neural Inf Process Syst* (2013) 26. doi:10.48550/arXiv.1310.4546

14. Zhang Y, Zhu Y, Liu J, Yu W, Jiang C. An interpretability optimization method for deep learning networks based on grad CAM. *IEEE Internet Things J* (2024) 12:3961–70. doi:10.1109/JIOT.2024.3485765

15. Oveis AH, Cantelli Forti A, Giusti E, Soltanpour M, Rojhani N, Martorella M. SHAP assisted resilience enhancement against adversarial perturbations in optical and SAR image classification. *IEEE Geosci Remote Sensing Lett* (2025) 22:1–5. doi:10.1109/LGRS.2025.3536005

16. Zafar MR, Khan N. Deterministic local interpretable model-agnostic explanations for stable explainability. *Machine Learn Knowledge Extraction* (2021) 3(3):525–41. doi:10.3390/make3030027

17. Luo X, Qin X, Wu Z, Yang F, Wang M, Shang J. Sediment classification of small-size seabed acoustic images using convolutional neural networks. *IEEE Access* (2019) 7:98331–9. doi:10.1109/ACCESS.2019.2927366

18. Chungath TT, Nambiar AM, Mittal A. Transfer learning and few-shot learning based deep neural network models for underwater sonar image classification with a few samples. *IEEE J Oceanic Eng* (2023) 49(1):294–310. doi:10.1109/JOE.2022.3221127

19. Jin L, Liang H, Yang C. Accurate underwater ATR in forward-looking sonar imagery using deep convolutional neural networks. *IEEE Access* (2019) 7:125522–31. doi:10.1109/ACCESS.2019.2939005

20. Cheng C, Wang C, Yang D, Wen X, Liu W, Zhang F. Underwater small target detection based on dynamic convolution and attention mechanism. *Front Mar Sci* (2024) 11:1348883. doi:10.3389/fmars.2024.1348883

21. Walker S, Peeples J, Dale J, Keller J, Zare A. Explainable systematic analysis for aperture sonar imagery. In: *2021 IEEE international geoscience and remote sensing symposium IGARSS*. IEEE (2021). p. 2835–8. doi:10.1109/IGARSS47720.2021.9554901

22. Richard GJ, Habonneau J, Guériot D, Le Caillec JM. AI explainability and acceptance: a case study for underwater mine hunting. *ACM J Data Inf Qual* (2024) 16(1):1–20. doi:10.1145/3635113

23. Keenan C, Miller MD. Human-aided explainable AI classification of forward-looking sonar imagery. In: *2024 IEEE/OES autonomous underwater vehicles symposium (AUV)*. IEEE (2024). p. 1–6. doi:10.1109/AUV61864.2024.11030772

24. Xie B, Zhang H, Wang W. Side-scan sonar image classification based on joint image deblurring-denoising and pre-trained feature fusion attention network. *Electronics* (2025) 14(7): 1287. doi:10.3390/electronics14071287

25. Zhu P, Zhang Y, Huang Y, Zhao C, Zhao K, Zhou F. Underwater Acoustic target recognition based on spectrum component analysis of ship radiated noise. *Appl Acoust* (2023) 211:109552. doi:10.1016/j.apacoust.2023.109552

26. Wang Z, Dahouda MK, Hwang H, Joe I. Explanatory LSTM-AE-Based anomaly detection for time series data in marine transportation. *IEEE Access* (2025) 13:23195–208. doi:10.1109/ACCESS.2025.3535695

27. Kubicek B, Sen Gupta A, Kirsteins I. Feature extraction and classification of simulated monostatic acoustic echoes from spherical targets of various materials using convolutional neural networks. *J Mar Sci Eng* (2023) 11(3):571. doi:10.3390/jmse11030571

28. Du X, Wen Y, Yan J, Zhang Y, Luo X, Guan X. Multi-target detection in underwater sensor networks based on Bayesian deep learning. *IEEE Trans Netw Sci Eng* (2025) 12(3):1581–96. doi:10.1109/TNSE.2025.3535572

29. Domingos LCF, Santos PE, Skelton PSM, Brinkworth RSA, Sammut K. A survey of underwater acoustic data classification methods using deep learning for shoreline surveillance. *Sensors* (2022) 22(6):2181. doi:10.3390/s22062181

30. Feng S, Ma S, Zhu X, Yan M. Artificial intelligence-based underwater acoustic target recognition: a survey. *Remote Sensing* (2024) 16(17):3333. doi:10.3390/rs16173333

31. Veitch E, Alsos OA. Human-centered explainable artificial intelligence for marine autonomous surface vehicles. *J Mar Sci Eng* (2021) 9(11):1227. doi:10.3390/jmse9111227

32. Qiao Y, Yin J, Wang W, Duarte F, Yang J, Ratti C. Survey of deep learning for autonomous surface vehicles in marine environments. *IEEE Trans Intell Transportation Syst* (2023) 24(4):3678–701. doi:10.1109/TITS.2023.3235911

33. Chen Z, Bao T, Zhang B, Wu T, Chu X, Zhou Z. Deep reinforcement learning methods for USV control: a review. In: *2023 china automation congress (CAC)*. IEEE (2023). p. 1526–31. doi:10.1109/CAC59555.2023.10450528

34. Yan J, You K, Cao W, Yang X, Guan X. Binocular vision-based motion planning of an AUV: a deep reinforcement learning approach. *IEEE Trans Intell Vehicles* (2023) 9:5299–315. doi:10.1109/TIV.2023.3321884

35. Liu H, Ye X, Zhou H, Huang H. Research on UUV carrying forward looking sonar for target location based on spatial analysis. *IEEE Trans Instrumentation Meas* (2025) 74:1–11. doi:10.1109/TIM.2025.3533664

36. Aubard M, Madureira A, Teixeira L, Pinto J. Sonar based deep learning in underwater robotics: overview, robustness, and challenges. *IEEE J Oceanic Eng* (2025) 50:1866–84. doi:10.1109/JOE.2025.3531933

37. Sun S, An W, Tian F, Nan F, Liu Q, Liu J, et al. A review of multimodal explainable artificial intelligence: past, present and future. *CoRR* (2024). doi:10.48550/arXiv.2412.14056

38. Li G, Wu M, Yang H. A new underwater acoustic signal recognition method: fusion of cepstral feature and multi path parallel joint neural network. *Appl Acoust* (2025) 239:110809. doi:10.1016/j.apacoust.2025.110809

39. Cai W, Zhu J, Zhang M. Multi-modality object detection with sonar and underwater camera *via* object-shadow feature generation and saliency information. *Expert Syst Appl* (2025) 128:128021. doi:10.1016/j.eswa.2025.128021

40. Zheng Z, Liu P. Underwater acoustic target recognition based on sub regional feature enhancement and multi activated channel aggregation. *J Mar Sci Eng* (2024) 12(11):1952. doi:10.3390/jmse12111952

41. Zhang Z, Wei C, Xie S, Zhang W, Wen L. A new multi sensor feature fusion kan network for autonomous underwater vehicle fault diagnosis. *IEEE Trans Instrumentation Meas* (2024) 74:1–11. doi:10.1109/TIM.2024.3522700