



OPEN ACCESS

EDITED BY

Taehwa Lee,
Toyota Motor North America, United States

REVIEWED BY

Mihailov Maria Emanuela,
Maritime Hydrographic Directorate, Romania
Ameen Noor,
Mustansiriyah University, Iraq

*CORRESPONDENCE

Lei Zhufeng,
✉ leizhufeng@xaau.edu.cn

RECEIVED 27 August 2025

ACCEPTED 20 October 2025

PUBLISHED 05 November 2025

CITATION

Muye Y, Zhufeng L, Yinuo D and Zhanrong S
(2025) A classification method of underwater
target radiated noise signals based on
enhanced images and convolutional neural
networks.
Front. Phys. 13:1693938.
doi: 10.3389/fphy.2025.1693938

COPYRIGHT

© 2025 Muye, Zhufeng, Yinuo and Zhanrong.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

A classification method of underwater target radiated noise signals based on enhanced images and convolutional neural networks

Yuan Muye¹, Lei Zhufeng^{2*}, Dang Yinuo¹ and Shi Zhanrong³

¹Department of Mechanical Engineering, Xihang University, Xi'an, China, ²National Joint Engineering Research Center for Special Pump Technology, Xihang University, Xi'an, China, ³HanCheng Development and Reform Commission, HanCheng, China

As the economy and society continue to develop, the range of underwater vehicles is expanding and technology is constantly being upgraded. Consequently, it is becoming increasingly difficult to classify and identify them, and the traditional classification method based on signal characteristics can no longer meet the urgent need for the accurate identification of underwater targets. This paper therefore proposes multiple convolutional neural network recognition methods based on enhanced Gramian Angular Field (GAF) images. Firstly, the radiated noise signals of underwater targets are converted into enhanced images using the GAF method. Then, the converted image dataset is used as input for the convolutional neural network. The input dataset is modified accordingly for each convolutional neural network. Finally, the significant advantages of convolutional neural networks in image processing are leveraged to achieve precise classification of underwater target radiated noise. In order to propose a convolutional neural network method that matches the enhanced image method, this paper compares the calculation results of multiple convolutional neural network models. The experimental results show that the VGG-16 model achieves greater classification accuracy and efficiency, reaching 80.67%.

KEYWORDS

underwater targets, radiated noise, enhanced images, convolutional neural network, Gramian angular field

1 Introduction

As the economy and society continue to develop, the variety of underwater targets is constantly increasing and technology is continuously improving. Furthermore, the non-stationary, nonlinear and non-Gaussian characteristics of radiated noise signals from underwater targets, coupled with multiple interacting factors, are making it increasingly difficult to classify these signals. Traditional methods of recognizing these signals mainly involve collecting them via sonar and processing them through time-series analysis before conducting classification and recognition through feature extraction. However, these methods are severely affected by environmental reverberation and the electrical noise inherent in the equipment, meaning they can no longer meet the

growing demand for data. Therefore, there is an urgent need for an advanced method that is less affected by environmental reverberation and equipment noise and offers improved classification accuracy and efficiency. The method described in this paper uses enhanced images based on Gramian Angular Field (GAF) and a convolutional neural network to classify underwater target radiated noise signals. No noise removal is performed in this process to ensure that no effective information in the signal is lost. The method is also less sensitive to environmental reverberation and equipment noise, thus improving the classification and recognition of underwater target radiated noise signals.

Extensive research has been conducted by scholars on traditional feature extraction methods for underwater target radiated noise signals. Li et al. [1] proposed a new nonlinear feature extraction method for ship radiated noise. This method reconstructs the phase space of the noise, maps time-domain data from low-dimensional phase space to high-dimensional phase space and inputs the extracted nonlinear features into a probabilistic neural network (PNN) classifier. The classification performance is quantified and the recognition rate is calculated. Compared with traditional feature extraction methods, the proposed nonlinear feature extraction method shows better separability and a higher recognition rate for different types of ship radiated noise. Du L et al. [2] proposed a joint feature extraction method for ship target recognition that combines energy features extracted by wavelet decomposition with frequency domain features extracted by a Mel filter. The recognition accuracy of joint features is significantly higher than that of single Mel frequency domain features, providing a useful reference for hydroacoustic target recognition. Wang et al. [3] proposed an improved finite rate of innovation (FRI) method for estimating the low-frequency line spectra of hydroacoustic passive target radiated noise under conditions of limited sampling points. This method can realise the high-precision extraction of line spectral information of vessel noise. Duan et al. [4–7] proposed a graph embedding method for time-domain ship radiated noise signals, as well as a fully parameterised prototype learning framework that can perform open-set identification of ship radiated noise signals end-to-end. Jin et al. [8, 9] introduced a multi-target feature extraction method based on endowment mode decomposition and statistical parameterised inverted surface coefficients. This method significantly increased the number of classifiable ship targets. Li GH et al. [10–13] proposed a quadratic decomposition and hybrid feature extraction method for hydroacoustic signals. The simulation results for measured ship radiated noise show that the proposed method achieves a recognition rate of 98.43%. However, traditional methods of extracting the radiated noise of underwater targets still face many challenges, such as heavy reliance on the operator's experience, low recognition accuracy and high computational costs, which severely hinder their development.

Scholars have conducted extensive research on machine learning for the detection of radiated noise from underwater targets. Ji et al. [14, 15] proposed a method combining image processing and deep self-coding networks to improve the low-frequency, weak line spectra of underwater targets in environments with a very low signal-to-noise ratio, achieving a line spectral density more than twice that extractable in the 10–300 Hz band. Gao et al. [16–18] proposed a two-stream deep learning network

with frequency characteristic transformation for line spectrum estimation to enhance the ability to learn target signal features. Xu et al. [19] proposed a new deep neural network model for underwater target recognition which integrates three-dimensional Meier frequency cepstral coefficients and three-dimensional Meier signs from ship audio signals as inputs. The method achieved an average recognition rate of 87.52% on the DeepShip dataset and 97.32% on the ShipsEar dataset, demonstrating strong classification performance. Ashok et al. [20–22] presented an innovative DNN called the Audio Perspective Region-based Convolutional Neural Network (APRCNN), which is now better able to identify, classify, and localise sounds in aquatic acoustic settings, as well as dealing with the unique problems of underwater signal processing. Yan CH et al. [23, 24] introduced a lightweight network based on a multi-scale asymmetric convolutional neural network (CNN) with an attention mechanism for ship radiated noise classification. Experiments on the DeepShip dataset showed that the recognition accuracy was 98.2%, with significantly fewer parameters. In recent years, machine learning-based methods for extracting underwater target radiated noise features have made great progress, and the extensive use of deep learning and graph methods has dramatically improved the accuracy and computational efficiency of underwater target radiated noise recognition, leading to the rapid development of these methods.

This paper proposes a method of classifying and recognizing underwater target radiated noise signals based on Gramian Angular Field (GAF)-enhanced images, utilizing convolutional neural networks. The original underwater target radiated noise signals are directly processed into GAF-based enhanced images, and the processed images are used as the input dataset for neural network training, and the great advantage of convolutional neural network in image processing is utilized to realize the high accuracy classification of underwater target radiated noise signals. In this paper, the optimal convolutional neural network model will be obtained by comparing multiple convolutional neural network methods to match the GAF enhanced image method.

This paper is structured as follows: [Section 2](#) outlines the overall architecture and computational flow of GAF-based enhanced image and convolutional neural network methods. [Section 3](#) details the mathematical foundations of different deep learning neural network classification methods. [Section 4](#) presents the experimental data, methodology and results. [Section 5](#) discusses and concludes.

2 Background

In this section, we provide a description of the relevant theoretical background, including the enhanced image conversion method and the CNN model.

2.1 GAF-based enhanced image conversion method

The steps to convert an underwater target radiated noise $A = x(t)$, ($t = 1, 2, \dots, N$) into an enhanced image are as follows:

Step 1: Let the underwater target radiated noise signal $X = x_1, x_2, \dots, x_N$ be normalized to the interval $[-1, +1]$. The normalization process is as follows:

$$x'(t) = \frac{(x(t) - \min(x))}{(\max(x) - \min(x))}$$

Step 2: The normalized underwater target radiated noise signal is converted into the Angle value in the polar coordinate system. For each x' to an Angle. The calculation process is as follows:

$$\theta(t) = 2\pi x'(t)$$

Step 3. Construct a $n \times n$ matrix, which is calculated from the cosine of the Angle difference between all time points. For each pair of time points i and j in the time series, the element $G(i, j)$ of the G matrix is calculated as follows.

$$G(i, j) = \cos(\theta(i) - \theta(j))$$

Step 4: Map the elements of G from $[-1, 1]$ to the range $[0, 1]$, and the final matrix G after mapping is goes as follows:

$$G = \begin{bmatrix} \cos(\varphi_{1,1}) & \cos(\varphi_{1,2}) & K & \cos(\varphi_{1,n}) \\ \cos(\varphi_{2,1}) & \cos(\varphi_{2,2}) & K & \cos(\varphi_{2,n}) \\ K & K & K & K \\ \cos(\varphi_{n,1}) & \cos(\varphi_{n,2}) & K & \cos(\varphi_{n,n}) \end{bmatrix}$$

Step 5: Each element $G(i, j)$ of the G matrix is converted into the corresponding gray value, and finally a gray image is obtained. The mapping formula is as follows:

$$Gray(i, j) = \left(\frac{G(i, j) + 1}{2} \right) \times 255$$

The matrix G captures the overall characteristics of the time series by converting time series information into angle information, and uses the Gramian matrix form to depict the global correlation between different time points in 2D space, thereby encoding the dynamic characteristics of one-dimensional signals into images. This approach not only uncovers the dynamic variations and evolving patterns across different time series but also captures the inherent structural characteristics within individual time series. To a certain extent, it clarifies the geometric connections and spatial configurations among local sequence segments, offering abundant feature expressions that empower models to more effectively grasp the underlying patterns within the sequences and enhance their learning performance. As a result, it functions as a valuable technique for extracting features from time series data, which will significantly contribute to our work.

The process of converting 1D time series into 2D images using the Gramian Angular Field (GAF) method is depicted in Figure 1. Figure 2 showcases the 2D images generated from five distinct underwater target radiated noise signals via the GAF method. From the distinct 2D images in Figure 2, it is evident that they possess unique visual features. These features encompass variations in texture, color, and local shapes, which highlight the specific attributes of each image. Texture differences may mirror the origin of the signal, while color variations can indicate the intensity or frequency of the signals. Furthermore, the local shapes present in these images can provide insights into the specific types of underwater target radiated noise signals, thereby further emphasizing the importance of analyzing these visual cues for effective classification tasks.

2.2 CNN

CNN is a powerful deep learning architecture that is widely used for image classification tasks. Its primary components include convolutional layers, pooling layers and fully connected layers. Convolution kernels are used in the convolutional layers to extract features from the input data. These layers are usually followed by pooling layers, which perform secondary feature extraction and reduce the dimensionality of the parameters to enhance computational efficiency. Following a series of convolution and pooling operations, the resulting feature maps are fed into the fully connected layers to facilitate the integration of all the neurons' outputs. The final classification is achieved using a Softmax layer. Successive convolution and pooling enable the CNN to progressively capture both low-level and high-level features from the input images, significantly reducing model training time and boosting computational performance.

Convolutional Layer: As a fundamental component of Convolutional Neural Networks (CNNs), this layer deploys multiple convolution kernels onto the input image. It extracts features by calculating the sum of element-wise multiplications between each kernel and local image regions, producing a 2D feature map. Each element in this map corresponds to the convolution kernel's response to the image at that specific location.

Activation Function: A nonlinear activation function is integrated to empower the network to learn and represent complex features, enabling it to model intricate patterns within the data.

Pooling Layer: This layer serves to reduce the spatial size of feature maps. A widely used technique is max-pooling: it partitions the feature map into non-overlapping subregions and selects the maximum value from each subregion as the pooled result, effectively downsizing the data while preserving critical information.

Fully Connected Layer: Following the convolutional and pooling layers, feature maps are flattened into a one-dimensional vector, which is then input into fully connected layers. In these layers, every neuron connects to all neurons in the preceding layer, learning weight parameters to map the extracted features onto different classes or target values.

Output Layer: For classification tasks, the output layer typically employs the softmax function to transform the output of the fully connected layer into a probability distribution across all classes. For regression tasks, it may consist of a linear layer or utilize alternative activation functions tailored to the regression objective.

Figure 3 shows the structure of CNN.

3 Proposed method

In this section, we present a classification framework that combines an enhanced image conversion method (GAF) with convolutional neural network for classifying original underwater target radiated noise signals.

The 1D time series is first converted into 2D images using the GAF method. Next, multiple convolution operations in the CNN can capture local features of the 2D images while preserving spatial

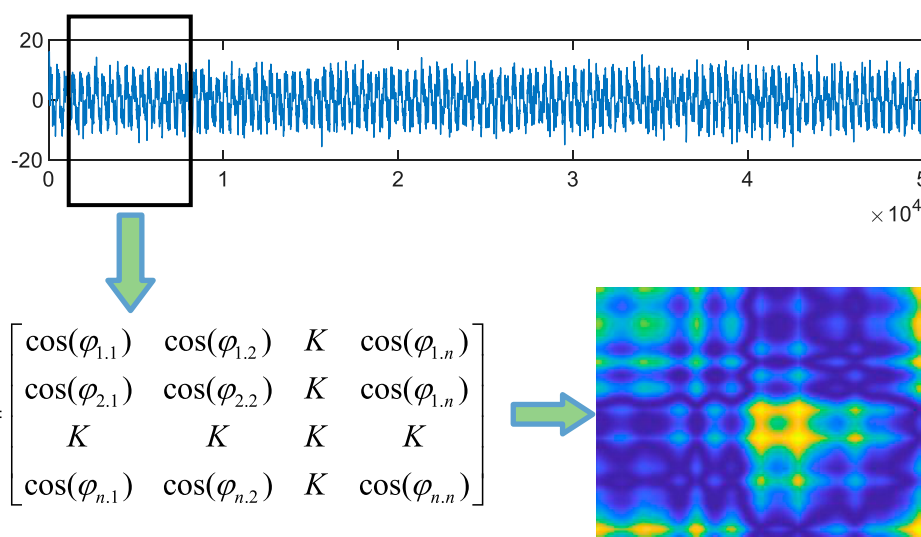


FIGURE 1
The conversion process of 1D time series into a 2D image using the GAF method.

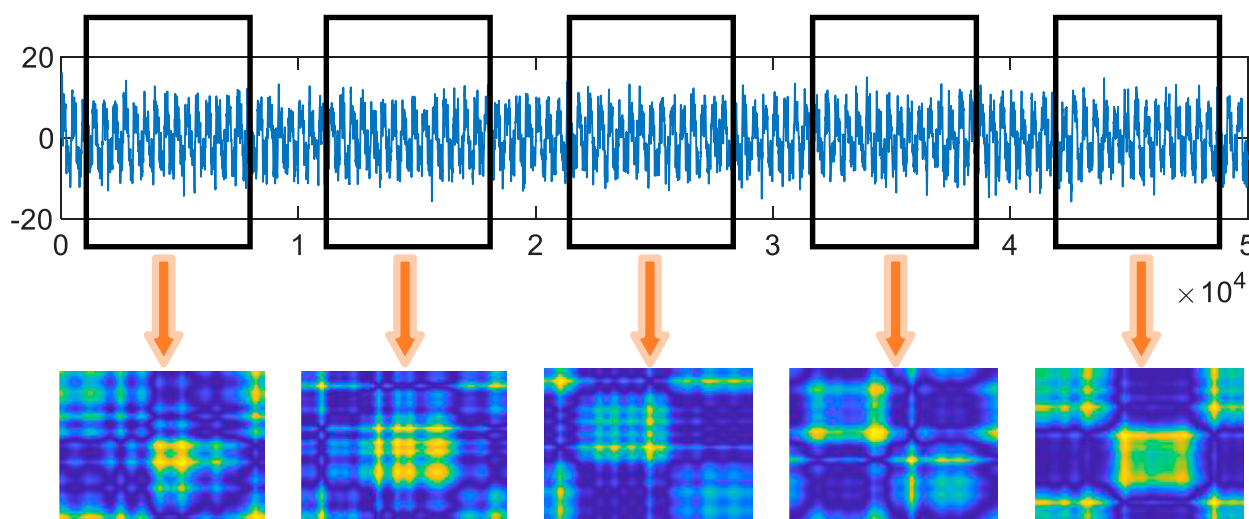


FIGURE 2
2D images generated from five different underwater target radiated noise signals using the GAF method.

structure information, and the pooling operations can lower the information volume while retaining important features. Features are fed into a fully connected network for feature recognition to achieve the final identification results.

Through comparing data of identification results, we can get the most suitable CNN to classify underwater target radiated noise signals.

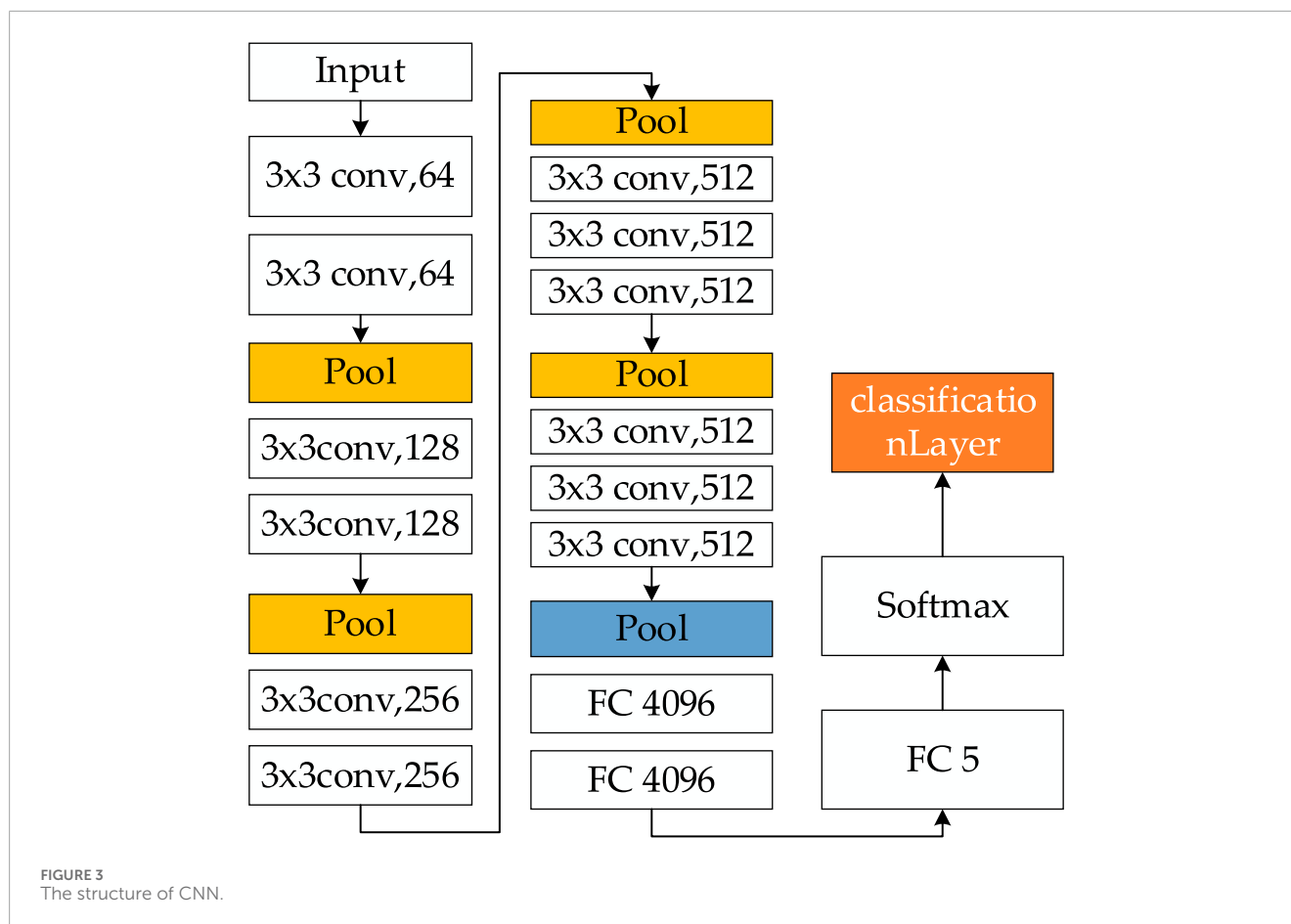
The proposed framework consists of the following steps:

Step 1: Raw underwater target radiated noise signals are transformed into enhanced image representations through GAF, without applying environmental noise suppression.

Step 2: The VGG-16 network is employed to extract discriminative features from the generated images and to perform classification. To ensure reliable model selection, a hold-out validation strategy is adopted, where 70% of the converted images were used to train the model, while the remaining 30% were used to validate it.

Step 3: During training, regularization techniques such as dropout are introduced to mitigate overfitting and enhance the generalization ability of the model.

Step 4: The classification outcomes are analyzed based on the independent test set to evaluate recognition performance and to validate the effectiveness and generalizability of



the proposed VGG-16-based framework for underwater acoustic signal processing.

The overall architecture of the proposed method is illustrated in Figure 4.

4 Experimental results and analysis

4.1 Experiments

In this study, the computing environment is a computer with a Windows 10 operating system and a single GPU. The GPU is an NVIDIA GeForce RTX 4070 SUPER.

To evaluate the effectiveness of proposed method, various CNN architectures are incorporated into the proposed framework, and their classification accuracies are compared. To further examine whether matching the proposed method with the most suitable CNN yields superior performance compared to traditional approaches, additional baseline models including LSTM and 1DCNN are also evaluated. The CNN architectures considered in this study include DarkNet-53, GoogLeNet, Inception-ResNet-v2, ResNet-18, ResNet-50, ResNet-101, ShuffleNet, SqueezeNet, VGG-19, and Xception. This comparative analysis aims to assess the effectiveness of different models in recognizing underwater acoustic signals. Experimental results demonstrate that the VGG-16 model outperforms all other

tested architectures, thereby confirming its suitability as the optimal backbone for the proposed framework.

4.2 Data introduction

The particularity of ship signal makes it difficult to obtain actual data, and the actual data has many interferences. The research goal is to verify the superiority of the algorithm rather than deal with complex interference noise, so we adopt analog signals. In this paper, the sampling time of the radiated noise signal from the underwater target adopted is 1 s, with a sampling frequency of 50 kHz.

In this study, different operating conditions were considered with variations in propeller blade number and displacement: Jiang4 represents a 4-blade propeller at 15 knots with a blade speed of 25 r/s and a displacement of 4,000 t; Jiang5 represents a 5-blade propeller at 15 knots with a blade speed of 25 r/s and a displacement of 2,000 t; and Jiang6 represents a 6-blade propeller at 15 knots with a blade speed of 25 r/s and a displacement of 2,000 t. In addition, to analyze the effect of ship speed under identical conditions, two cases were defined as speed10 and speed20, both with a 4-blade propeller, a blade speed of 25 r/s, and a displacement of 2,000 t, corresponding to ship speeds of 10 knots and 20 knots, respectively. The data parameters are shown in Table 1 and an example of the underwater target radiated noise signal is shown in Figure 5.

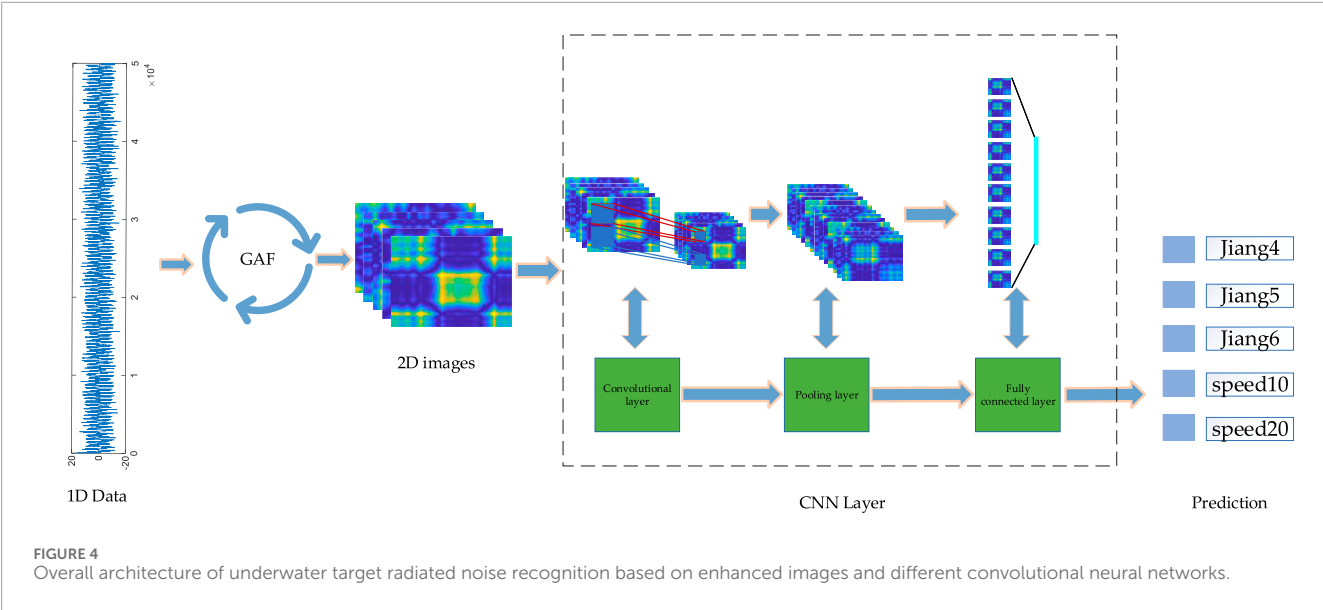
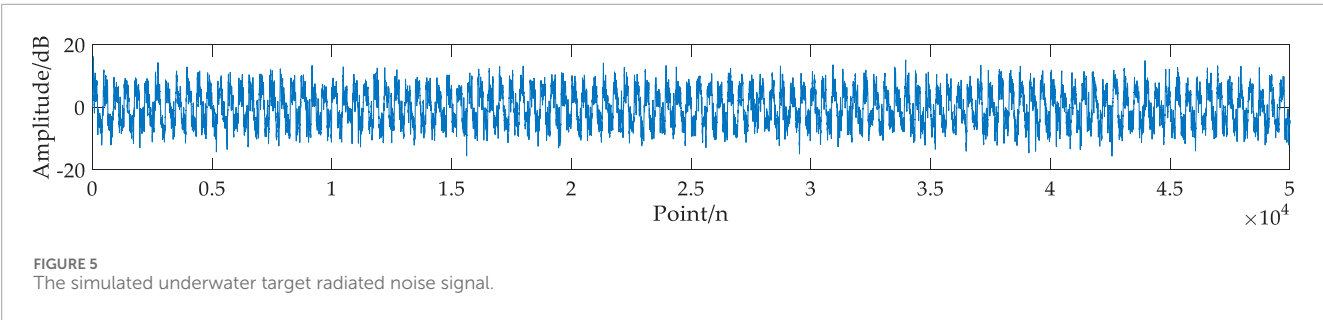


TABLE 1 The parameters of underwater target radiated noise signal simulation.

Types	Number of paddles	Speed (kn)	Paddle speed(r/s)	Sampling time(s)
jiang4	4	15	25	1
Jiang5	5	15	25	1
Jiang6	6	15	25	1
speed10	4	10	25	1
speed20	4	20	25	1



4.3 Determination of network parameters

This section outlines the determination of network parameters for the convolutional neural network (CNN).

Step 1: The input 2D images were resized and standardized to conform to the dimensional requirements of the employed CNN architectures, thereby ensuring consistency across all models during training and evaluation. The image size generated by the enhanced image conversion method is [224, 224, 3]. To maintain compatibility with different network architectures, the input size is adjusted accordingly

without reducing the original resolution. This strategy helps retain the complete information of the input image, avoiding key feature loss caused by image cropping or resizing, and thereby facilitating more effective feature extraction and classification.

Step 2: Define the structural components of the network, including the type and number of convolutional layers, activation functions, pooling layers, fully connected layers, and the output layer. Alternatively, pre-trained CNN architectures can be adopted and adapted to the specific task.

TABLE 2 Convolutional neural network structure parameters.

Optimiser	MaxEpochs	LearnRate	Mini batch size	Validation frequency
sgdm	50	0.0001	15	10

Step 3: Configure and fine-tune the key network parameters, such as the number of filters, kernel size, stride, padding, activation function type, dropout rate, learning rate, and batch size.

In the training of convolutional neural networks, the selection of hyperparameters plays a crucial role in determining both the convergence behavior and the generalization performance of the model. Factors such as the optimizer, learning rate, maximum number of epochs, mini-batch size, and validation frequency not only influence training efficiency but also directly affect the model's accuracy on unseen data. In this study, the stochastic gradient descent with momentum (SGDM) optimizer is employed, as it has been widely demonstrated in image recognition tasks to strike a favorable balance between convergence speed and generalization capability. Compared with adaptive optimization methods (e.g., Adam), SGDM often yields superior test accuracy. The maximum number of epochs (MaxEpochs = 50) is set to ensure that the model undergoes sufficient training to capture discriminative features, while avoiding excessive computational cost and the risk of overfitting. To maintain stable parameter updates, a relatively small learning rate (LearnRate = 0.0001) is adopted, which enables the network to gradually approach the optimal solution in a complex loss landscape without oscillation or divergence. Furthermore, a mini-batch size of 15 is chosen to achieve a balance between gradient variance and computational efficiency; smaller batches introduce beneficial stochasticity in gradient estimation, which facilitates escaping local minima and enhances generalization. Finally, the validation frequency (Validation Frequency = 10) is set to monitor model performance at appropriate intervals during training, allowing timely adjustments to suppress potential overfitting. Based on these considerations, the hyperparameter configuration summarized below is adopted in the subsequent experiments.

The detailed configuration used in this study is presented in Table 2.

4.4 Result

In this study, 70% of the converted images were used to train the model, while the remaining 30% were used to validate it. The computing environment is a computer with a Windows 10 operating system and a single GPU. The GPU is an NVIDIA GeForce RTX 4070 SUPER.

The results of the experiments demonstrate that the VGG-16 model achieves the best performance of all the models tested, indicating its suitability as the optimal model structure for the proposed framework and better than sequential models. Figure 6 shows the classification results. Table 3 showcase the data directly.

4.5 Analysis of different convolutional neural networks

As demonstrated in Figure 6 and Table 3, the classification accuracy of VGG-16 is markedly higher than that of other convolutional neural network architectures. The rationale for selecting VGG-16 as the baseline model lies in its architectural simplicity and effectiveness. Its design strategy of stacking multiple 3×3 convolutional layers allows the network to achieve substantial depth, thereby enhancing its capacity to learn complex and hierarchical feature representations while maintaining a manageable number of parameters. Moreover, the modular and uniform nature of the VGG-16 design makes it highly adaptable and transferable across a broad range of image classification tasks, which further justifies its selection for this study.

Several architectural features contribute directly to the superior performance of VGG-16. First, the convolutional layers are each followed by ReLU activation functions, which introduce nonlinearity and enable the network to model complex patterns. In the absence of such activation functions, stacking multiple convolutional layers would reduce to a linear transformation, severely limiting the representational capacity of the model. Second, the inclusion of three fully connected layers provides multiple levels of abstraction, with deeper layers capturing increasingly complex and higher-order features. This design enables the integration of information from multiple dimensions, thereby improving classification performance. Third, the consistent use of 3×3 convolution kernels increases the effective receptive field while preserving fine-grained details and contextual information, which enriches feature representations and enhances the expressive power of the model. Finally, the application of 2×2 max pooling layers with a stride of 2 ensures that the most salient local features are retained, strengthening the discriminative ability of the extracted features.

When compared to alternative architectures, the superiority of VGG-16 becomes more evident. DarkNet-53, while powerful for large-scale object detection, is considerably deeper and prone to overfitting when trained on relatively small datasets such as underwater acoustic signals. GoogLeNet adopts an Inception module structure, which increases architectural complexity but may dilute local feature extraction, thereby reducing its effectiveness in scenarios requiring fine-grained classification. Inception-ResNet-v2 integrates residual connections with Inception modules, but its depth and parameter complexity can hinder convergence and generalization under limited data conditions. ResNet-18/50/101 benefit from residual learning, yet their skip connections may lead to an overemphasis on global features at the expense of local detail, which is critical in underwater target noise classification. ShuffleNet and SqueezeNet prioritize computational efficiency and lightweight design, but their reduced parameterization limits representational capacity, resulting in lower accuracy. VGG-19, though structurally similar to VGG-16, introduces additional convolutional layers,

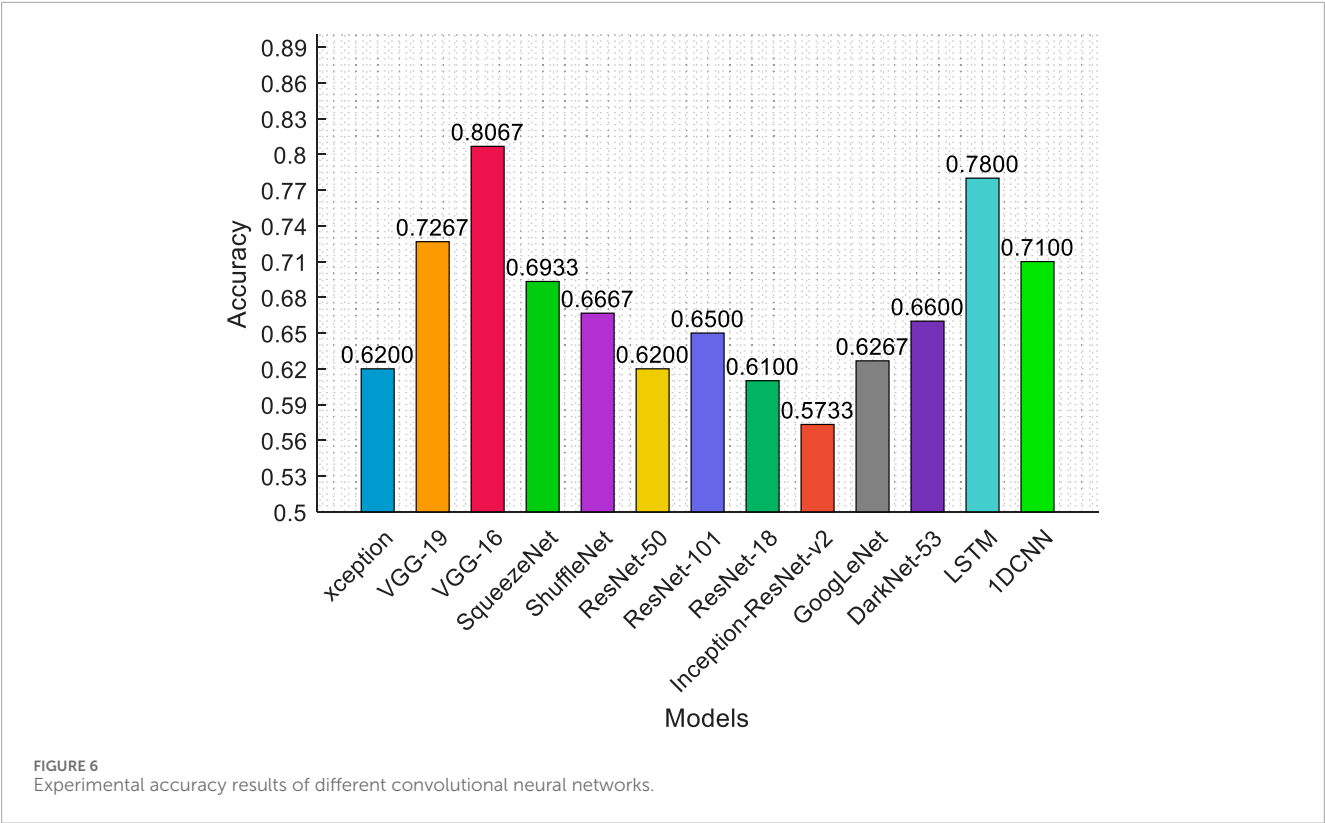


TABLE 3 Image classification results transformed by multiple convolutional neural networks.

Xception	VGG-19	VGG-16	Inception-ResNet-v2	GoogLeNet
0.62	0.7267	0.8067	0.5733	0.6267
SqueezeNet	ShuffleNet	ResNet-50	ResNet-101	LSTM
0.6933	0.6667	0.62	0.65	0.78
1DCNN	ResNet-18	DarkNet-53		
0.71	0.61	0.6267		

which increase computational cost without yielding significant performance gains in this task. Finally, Xception employs depthwise separable convolutions, which are effective in large-scale image recognition but tend to underperform on domain-specific datasets where local and low-level features dominate.

Beyond CNN variants, VGG-16 also demonstrates clear advantages over sequential models such as LSTM and shallow architectures such as 1D-CNN. LSTM networks, while effective for capturing long-term temporal dependencies, are less adept at extracting local spatial correlations when one-dimensional signals are transformed into 2D representations. This limits their ability to fully exploit fine-grained structural patterns inherent in acoustic spectrograms. Similarly, 1D-CNN models, although computationally efficient, operate on raw one-dimensional sequences and thus fail to capture the rich 2D feature hierarchies available in time–frequency representations. In contrast, VGG-16 leverages its hierarchical convolutional structure to capture both

local and global feature dependencies within spectrograms, enabling superior discrimination of subtle acoustic variations.

In summary, VGG-16 achieves an optimal balance between architectural depth, feature representation capacity, and training stability. By virtue of its simple yet powerful design, it demonstrates superior capability in local feature extraction, stronger nonlinear modeling capacity, and more stable convergence. Experimental results further confirm the robustness and adaptability of VGG-16 in classifying underwater target radiated noise signals, clearly outperforming more complex or lightweight alternatives, as well as sequential architectures such as LSTM and 1D-CNN.

As shown in Table 4, although models such as ResNet and Inception-ResNet-v2 are powerful for large-scale datasets, their excessive depth and architectural complexity often lead to overfitting and unstable convergence when applied to relatively small and domain-specific datasets such as underwater acoustic signals. Conversely, lightweight networks such as

TABLE 4 Advantages and disadvantages of different CNNs.

CNN architecture	Key characteristics	Advantages	Limitations in this study
VGG-16	16 layers; stacked 3×3 conv kernels; 3 FC layers	Strong local feature extraction; balanced depth; stable convergence	Slightly higher computational cost than lightweight models, but acceptable
VGG-19	Deeper variant of VGG-16 (19 layers)	Similar feature representation as VGG-16	Higher computational burden without notable accuracy gain
ResNet-18/50/101	Residual skip connections	Alleviates vanishing gradient; effective for very deep networks	May overemphasize global features, weaker local detail extraction for acoustic signals
GoogLeNet	Inception modules with multi-scale filters	Efficient parameter usage; multi-scale representation	Complexity may dilute fine-grained local features; less effective for small datasets
Inception-ResNet-v2	Inception modules + residuals; very deep	Strong representation power in large-scale tasks	Overly complex; prone to overfitting on limited acoustic data
DarkNet-53	Deep network (53 layers); backbone for YOLOv3	High performance in object detection	Excessive depth; requires large-scale data, less stable on small datasets
Xception	Depthwise separable convolutions	Computational efficiency; good for large-scale images	Less effective in capturing domain-specific local features
ShuffleNet	Channel shuffle for lightweight design	Very fast; low resource consumption	Limited representational capacity; accuracy loss on complex tasks
SqueezeNet	Fire modules; very small parameter count	Extremely lightweight	Sacrifices accuracy for compactness; weak performance in feature-rich tasks

ShuffleNet and SqueezeNet are computationally efficient but lack sufficient representational power to capture subtle acoustic features. In contrast, VGG-16 achieves an optimal trade-off between depth, feature extraction capacity, and training stability, making it particularly well-suited for this classification task.

5 Experimental results and analysis

This study introduces a novel approach for time-series data representation by employing the Gramian Angular Field (GAF) method to transform 1D underwater acoustic signals into 2D images. Building upon this representation, the proposed framework leverages convolutional neural networks (CNNs) for feature extraction and classification of underwater target radiated noise signals. Among the tested architectures, the integration of VGG-16 into the proposed model yields the highest classification accuracy, reaching 80.67%, thereby demonstrating its superiority over alternative CNNs, as well as traditional sequence modeling approaches such as LSTM and 1D-CNN. The strong performance of VGG-16 highlights its effectiveness in capturing local, fine-grained features that are critical for distinguishing subtle acoustic patterns in underwater environments.

Although the method proposed in this paper achieves good results in the classification of underwater target radiated noise signals, it inevitably has certain limitations. First, the proposed method involves numerous hyperparameters, including the number of hidden layer nodes, the L2 regularization coefficient, and the initial learning rate. During model training, it is necessary to precisely determine the optimal values of these hyperparameters simultaneously. Additionally, the proposed model has significant complexity, which inevitably increases the computational cost during model training. This not only imposes higher demands on hardware infrastructure but may also limit the model's scalability and deployment in practical applications. Therefore, reducing model complexity while maintaining accuracy has become an important research direction. To address these limitations, we offer several suggestions. For hyperparameter tuning, we can utilize effective global optimization techniques to simultaneously optimize all model hyperparameters with fewer experimental trials, thereby improving training efficiency and overall model performance. In addressing model complexity, pruning and lightweight techniques can be explored to reduce model parameter size and optimize model structure. By implementing these methods, we can more effectively address challenges related to hyperparameter tuning and model complexity, thereby promoting the continued advancement of this research.

From a holistic perspective, another limitation lies in the fact that the particularity of ship signals makes it difficult to obtain actual data. Therefore, in the experimental part of this paper, simulated data was used. However, in the real marine environment, more complex interferences and environmental noises are often encountered, such as multipath propagation, non-Gaussian background noise, and multi-target interference, etc. These factors may affect the performance of the proposed method. Thus, in subsequent research, we will conduct experiments and validations in real marine scenarios to comprehensively evaluate the practicality and robustness of the method.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

YM: Writing – original draft. LZ: Funding acquisition, Writing – original draft, Writing – review and editing. DY: Methodology, Writing – review and editing. SZ: Supervision, Writing – review and editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

References

- Li Y, Shen Y. A novel nonlinear feature extraction method of ship radiated noise based on phase space reconstruction and multiscale improved permutation entropy*. In: *2024 21st International bhurban conference on applied sciences and technology* (2024). p. 672–7.
- Du L, Liu M, Lv Z, Wang Z, Wang L, Wang G. Classification method for ship-radiated noise based on joint feature extraction. In: C Cruz, Y Zhang, W Gao, editors. *Intelligent computers, algorithms, and applications: third BenchCouncil international symposium, IC 2023, revised selected papers. Communications in computer and information science* (2024). p. 75–90.
- Wang X, Guo L, Sun H. Frequency estimation of underwater acoustic targets based on improved FRI method. *J Phys Conf Ser* (2024) 2718:012102. doi:10.1088/1742-6596/2718/1/012102
- Duan Y, Shen X, Wang H. Time-domain anti-interference method for ship radiated noise signal. *Eurasip J Adv Signal Process* (2022). doi:10.1186/s13634-022-00895-y
- Duan Y, Shen X, Wang H. Dual-branch time-frequency domain anti-interference method for ship radiated noise signal. *Ocean Eng* (2023) 279:114537. doi:10.1016/j.oceaneng.2023.114537
- Duan Y, Shen X, Wang H, Yan Y. Module-level soft fault detection method for typical underwater acoustic sensing system. *Sensors and Actuators A-Physical* (2023) 354:114274. doi:10.1016/j.sna.2023.114274
- Duan Y, Shen X, Wang H, Yan Y. An open-set recognition method for ship radiated noise signal based on graph convolutional neural network prototype learning. *Digital Signal Process*. (2025) 156:104748. doi:10.1016/j.dsp.2024.104748
- Jin S, Su Y, Guo C, Fan Y, Tao Z. Offshore ship recognition based on center frequency projection improved EMD and KNN algorithm. *Mech Syst Signal Process* (2023) 189. doi:10.1016/j.ymssp.2022.110076
- Jin S, Su Y, Guo C, Ma C, Fan Y, Tao Z. Intrinsic mode ensemble statistical cepstral coefficients for feature extraction of ship-radiated noise. *Appl Acoust* (2025) 227:110255. doi:10.1016/j.apacoust.2024.110255
- Li G, Han Y, Yang H. A novel approach for underwater acoustic signal denoising based on improved time-variant filtered empirical mode decomposition and weighted fusion filtering. *Ocean Eng* (2024) 313:119550. doi:10.1016/j.oceaneng.2024.119550
- Li G, Liu B, Yang H. Research on feature extraction method for underwater acoustic signal using secondary decomposition. *Ocean Eng* (2024) 306:117974. doi:10.1016/j.oceaneng.2024.117974
- Li G, Yan H, Yang H. Secondary decomposition multilevel denoising method of hydro-acoustic signal based on information gain fusion feature. *Nonlinear Dyn* (2025) 113:5251–89. doi:10.1007/s11071-024-10539-5
- Yang H, Lai M, Li G. Novel underwater acoustic signal denoising: combined optimization secondary decomposition coupled with original component processing algorithms. *Chaos Solitons Fractals* (2025) 193:116098. doi:10.1016/j.chaos.2025.116098
- Ji F, Li G, Lu S, Ni J. Research on a feature enhancement extraction method for underwater targets based on deep autoencoder networks. *Appl Sciences-Basel* (2024) 14:1341. doi:10.3390/app14041341
- Ji F, Ni J, Li G, Liu L, Wang Y. Underwater Acoustic target recognition based on deep residual attention convolutional neural network. *J Mar Sci Eng* (2023) 11:1626. doi:10.3390/jmse11081626
- Gao W, Liu Y, Chen D. A dual-stream deep learning-based Acoustic denoising model to enhance underwater information perception. *Remote Sensing* (2024) 16:3325. doi:10.3390/rs16173325

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 author(s) 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.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphy.2025.1693938/full#supplementary-material>

17. Gao W, Liu Y, Zeng Y, Liu Q, Li Q. SAR image ship target detection adversarial attack and defence generalization research. *Sensors* (2023) 23:2266. doi:10.3390/s23042266
18. Gao W, Lv Y, Li X, Yu G, Wang H. An integrated calculation model for vibroacoustic radiation and propagation from underwater structures in Oceanic acoustic environments. *Ocean Eng* (2024) 304:117682. doi:10.1016/j.oceaneng.2024.117682
19. Xu W, Han X, Zhao Y, Wang L, Jia C, Feng S, et al. Research on underwater acoustic target recognition based on a 3D fusion feature joint neural network. *J Mar Sci Eng* (2024) 12:2063. doi:10.3390/jmse12112063
20. Ashok P, Latha B. Absorption of echo signal for underwater acoustic signal target system using hybrid of ensemble empirical mode with machine learning techniques. *Multimedia Tools Appl* (2023) 82:47291–311. doi:10.1007/s11042-023-15543-2
21. Ashok P, Latha B. Develop an improved mel-frequency cepstral coefficients signal processing algorithms for enhancing underwater acoustic signal through wireless network. *Measurement* (2025) 243:116414. doi:10.1016/j.measurement.2024.116414
22. Ashok P, Latha B. Feature extraction of underwater acoustic signal target using machine learning technique. *Traitement Du Signal* (2024) 41:1303–14. doi:10.18280/ts.410319
23. Yan C, Yan S, Yao T, Yu Y, Pan G, Liu L, et al. A lightweight network based on multi-scale asymmetric convolutional neural networks with attention mechanism for ship-radiated noise classification. *J Mar Sci Eng* (2024) 12:130. doi:10.3390/jmse12010130
24. Yan C, Yao T, Yu Y, Yang C, Liu L, Pan G, et al. Underwater target recognition using a lightweight asymmetric convolutional neural network. In: *17th acm international conference on underwater networks and systems, wuwnet 2023* (2024).