VOL. 25, NO. 1, 2026, 14-21
elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

Feature-Enriched Deep Learning: CAPSE-Based Extraction with ResNet50 for Underwater Acoustic Classification

Najamuddin[1], **Usman Ullah Sheikh**[1*] and **Ahmad Zuri Sha'ameri**[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.

[*]Corresponding authors: usman@fke.utm.my

Abstract: Passive acoustic classification acts as a major function in automated ship identification, where deep learning models are used to recognize ship types from radiated noise. A major challenge lies in embedding domain-specific knowledge into these models to improve feature discrimination. This study proposes a classifier that integrates Coherently Averaged Power Spectral Estimation (CAPSE) method with a ResNet50-based classifier. Initially, ships' frequency spectrums are processed using CAPSE analysis, enabling the extraction of key machinery characteristics, and are transformed into LOFAR grams. These time-frequency representations are subsequently processed by a ResNet50 network, which leverages deep convolutional architectures to capture hierarchical feature representations. Taking advantage of transfer learning with ResNet50 deep feature extraction capabilities, the proposed model effectively learns complex patterns within the data, leading to improved classification performance. Evaluation on the standard DeepShip dataset showed that the proposed methodology attained an average classification accuracy and F1-score of 89.93%, while maintaining good generalization and robustness. tSNE feature visualization demonstrated improved class separation by the trained model, with most samples correctly clustered. This study improves underwater acoustic classification by showing that combining deep learning with better feature extraction can lead to highly accurate results.

Keywords: Passive acoustic classification, CAPSE, ResNet50, CNN, LOFAR gram

Article History: received 27 March 2025; accepted 17 June 2025; published 30 April 2026
Digital Object Identifier 10.11113/elektrika.v25n1.716

1. INTRODUCTION

Marine ship classification based on their acoustic signatures generated by onboard machinery is critical for surveillance, observation, and security in underwater applications [1]–[3]. Accurate discrimination among marine ship types is critical for both defense-related tasks, including the identification of potentially hostile targets, and civilian applications such as maritime traffic monitoring and management [4]. Unlike radar or visual methods, which use electromagnetic waves, acoustic systems enable detection in long-range and stealth conditions [5]. However, extracting meaningful frequency features is challenging due to ships' operational variations and ambient noise. Lower signal-to-noise ratios of the ships' signals further complicate the classification, with studies showing that ships become undetectable under -14.4 dB, and 90% experiencing SNR below 0 dB in ocean noise [6]. The underwater acoustic environment is highly challenging because of intricate propagation phenomena, pervasive ambient noise, and interference from multiple concurrent sources, which together make the problem especially demanding [7]. Conventional classification approaches perform well with supervised settings but struggle in real-world underwater environments due to high noise levels and uncertainty [8]. This highlights the need to incorporate advanced techniques that enhance the ship's machinery signature and leverage deep learning for more accurate and reliable classification.

Conventional methods for underwater acoustic target classification utilize signal processing approaches like Mel-frequency cepstral coefficients, wavelet analysis, and Fourier transforms [9]–[11]. These techniques excel at extracting distinctive features from noise-free signals [12]. However, these methods perform poorly when faced with severely noisy or heavily altered signals. With the progression of artificial intelligence, Convolutional Neural Networks (CNN) have been applied for acoustic signal classification by converting signals into spectrograms, allowing the task to be approached as an image recognition problem [13]. CNNs for underwater acoustic classification often struggle with limited training data, increased computational demands, and difficulties in capturing complex acoustic patterns[14]. These models require extensive tuning and large datasets to generalize well, making them less practical for real-world underwater applications. Fine-tuning ResNet50 offers a more efficient alternative by leveraging pre-trained weights from large-scale datasets, changing only the lower layers to extract features [15]. This approach speeds up convergence, improves generalization with limited data, and reduces computational cost compared to training from scratch. By

preserving lower-layer features that capture deep low-level patterns, fine-tuning enhances classification performance in noisy underwater environments while requiring fewer training resources. A major obstacle in underwater acoustics is the lower signal levels of acoustic sources due to ambient noise, surface reflections, and numerous forms of interference [16]. By integrating power spectra over multiple measurements, preprocessing algorithms such as Coherently Averaged Power Spectral Estimation (CAPSE) enhance acoustic signal quality by suppressing noise and boosting the visibility of key spectral features [17].

This study is built on the integration of CAPSE and ResNet50. CAPSE is employed as a preprocessing method to enhance the tonal components of acoustic signals by suppressing noise and reinforcing target-related features using coherent spectral averaging. The resulting signals are transformed into LOFAR gram images and subsequently input to a modified pretrained network to extract weak machinery-related features. The formulated method was tested on a publicly available dataset, DeepShip [18], delivering higher accuracy and enhanced generalization in noisy conditions. An overview in Figure 1 is included to illustrate the classification pipeline and contextualize the challenge. This approach highlights the need for an accurate preprocessing method and pre-trained models to improve classification operations in a noisy environment.

The rest of the article is arranged as follows. Section 2 surveys relevant research in the relevant research area. Section 3 describes the proposed methodology, including acoustic signal preprocessing, dataset construction, and the experimental model parameters. Section 4 presents and analyzes the results, highlighting the benefits of the proposed approach. Finally, Section 5 concludes the paper by reviewing the core contributions and outlining directions for future work.

Figure 1. Overview of the classification pipeline showing acoustic input, preprocessing, and vessel classifier

2. RELATED WORK

Studies on marine ship identification based on radiated acoustic noise have explored a wide selection of signal processing methods and machine learning models. Earlier attempts primarily relied on manual analysis of acoustic signatures by sonar operators, and classification performance is dependent on their expertise [19]. However, recent developments in computational capabilities and deep learning strategies have significantly accelerated progress in this area. There has been a growing interest in automated classification.

One of the most widely used preprocessing techniques is the Fast Fourier Transform (FFT), which transforms time-domain signals into the frequency domain, enabling the identification of spectral components to analyze inherent patterns [20]. The wavelet transform provides both time-frequency resolution for analyzing non-stationary signals, enabling the detection of transient features across multiple time scales [21]. Mel-Frequency Cepstral Coefficients (MFCCs) are widely employed in audio analysis due to their effectiveness in capturing the perceptual characteristics of sound signals [22]. MFCCs are designed to replicate the auditory perception sensitivity to frequency variations, making them particularly effective for speech and sound classification tasks [23]. The Constant-Q Transform offers variable resolution across the frequency spectrum, providing a more flexible approach to acoustic analysis [24]. Beyond these general techniques, LOFAR (Low-Frequency Analysis and Recording) focuses on identifying prolonged spectral patterns, making it particularly effective for detecting deterministic components like engine and mechanical noises. In contrast, the DEMON (Detection of Envelope Modulation on Noise) method targets the extraction of modulation frequencies produced by rotating elements, such as propellers [25].

Marine ship classification has been investigated using a variety of multimodal recognition strategies. In [26]The approach combined visual imagery with radiated acoustic noise to provide complementary information for classification. In [27], another study adopted a multi-chunk spectral investigation framework to represent diverse sub-band frequency characteristics, thereby offering a richer description of the acoustic scene. In a related effort, underwater target noise classification was enhanced through the joint extraction of one-third octave spectra, PSD, and MFCC features [28]. Collectively, these methods seek to increase classification performance by fusing heterogeneous feature representations and strengthening the strength of the recognition framework.

Machine learning approaches, including Support Vector Machines (SVM) and Artificial Neural Networks (ANN), have been applied to acoustic classification tasks. These methods depend on feature extraction, where real-world recorded acoustic signals are transformed into feature sets for the classifier [29]. For instance, Sherin and Supriya [30], utilized enhanced SVM classifiers to differentiate ship noise types. They used MFCC features as inputs to a neural network adjusted through the Salp Swarm Algorithm [31]. However, these networks face challenges in capturing deep, complex features due to their minimal design. To get around these limitations, CNNs have been employed to directly map raw waveforms or time-frequency representations to ship classes [27], [32]. CNNs have demonstrated strong performance in ship classification using acoustic signals. For example, Xu Cao et al. [33], introduced a CNN combining Second-Order Pooling (SOP) and the Constant-Q Transform for feature processing, surpassing conventional classifiers such as Deep Belief Networks and VGG-Net. Similarly, a custom CNN architecture has also been proposed to improve LOFAR spectrogram classification. Lucas Cinelli et al.

[34] suggested VesselNet for spectrogram-based ships classification, utilizing the Two-Pass Split-Window filter [35] to enhance feature extraction.

This paper proposes a pre-trained ResNet50 architecture, with fine-tuning, as a classifier. This approach to architecture significantly lowers the dependency on large datasets and resource requirements.

3. METHODOLOGY

This segment presents the approach for passive acoustic target classification, combining CAPSE for signal feature enhancement with a pretrained ResNet50 model for classification. CAPSE increases spectral clarity by reducing the noise floor, while ResNet50 extracts deep features from the refined LOFAR grams to identify patterns, thereby enhancing classification accuracy.

3.1 CAPSE

CAPSE is an advanced signal processing method devised to improve the recognition of sinusoidal tonals in noisy signals. In contrast to conventional approaches such as the Welch's method periodogram, CAPSE preserves phase consistency through several segments of the signal, leading to a notable improvement in signal-to-noise ratio (SNR) [36].

For a sinusoidal signal distorted by the presence of noise, the Fourier transform for every chunk is represented as:

$$S_k(\omega) = S_0(\omega)e^{(j\phi_k)} \quad (1)$$

where $\phi_k = \omega_0 kD$ denotes the phase change linking the Fourier transforms of the k^{th} and the first segments at frequency ω_0. CAPSE performs coherent averaging across multiple segments to improve the signal-to-noise ratio:

$$\overline{X}(\omega) = (1/K)\Sigma_{(k=0)}^{(K-1)}X_k(\omega)e^{(-j\phi_k)} \quad (2)$$

This offset causes a phase deviation within the chunks, which is fixed by applying another DFT along the chunk indices as shown in (3), resulting in:

$$\hat{X}(\omega_l, \omega_m) = (1/K)\Sigma_{(k=0)}^{(K-1)}X_k(\omega_l)e^{(-j\omega_v k)} \quad (3)$$

The CAPSE spectra is then characterized as:

$$P_{x_{xx}}^{CAPSE}(\omega) = (1/UM)\left|\hat{X}((\omega_l, \omega_{\delta l})\right|^2 \quad (4)$$

By emphasizing the dominant energy components, CAPSE retains the maximum critical spectral data, making it a robust approach for detecting sinusoidal signals in noisy surroundings. Further information on the process can be found in [36]. Figure 2 shows three samples of LOFAR grams created by periodogram, Welch, and CAPSE methods. The tonal components estimated using CAPSE exhibit greater clarity compared to the other two methods.

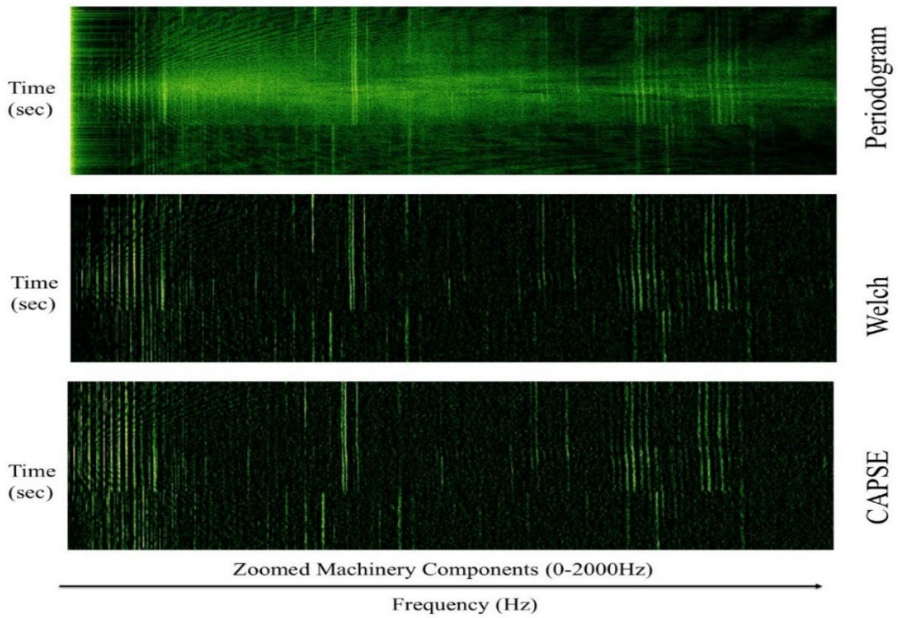

Figure 2. Sample LOFAR grams created via periodogram, Welch and CAPSE respectively

3.2 ResNet50 classifier

In this research work, we propose a pre-trained CNN model. However, CNN architecture has certain limitations, particularly in terms of computational expense, which increases with network depth and the number of parameters. Training and inference with deep CNNs require extensive computational resources. To address this challenge, ResNet50 is selected, as its residual connections facilitate a more efficient training process, enabling faster convergence and reducing overall computational complexity [37]. We have used the ResNet-50 model available in the high-level neural network API, MATLAB® 2024a, within the Deep Network Designer framework.

The ResNet-50 model in MATLAB was used with pre-trained weights, originally trained to recognize 1,000 object classes on the ImageNet dataset. To adapt this architecture for classifying 4 ship classes, the final densely

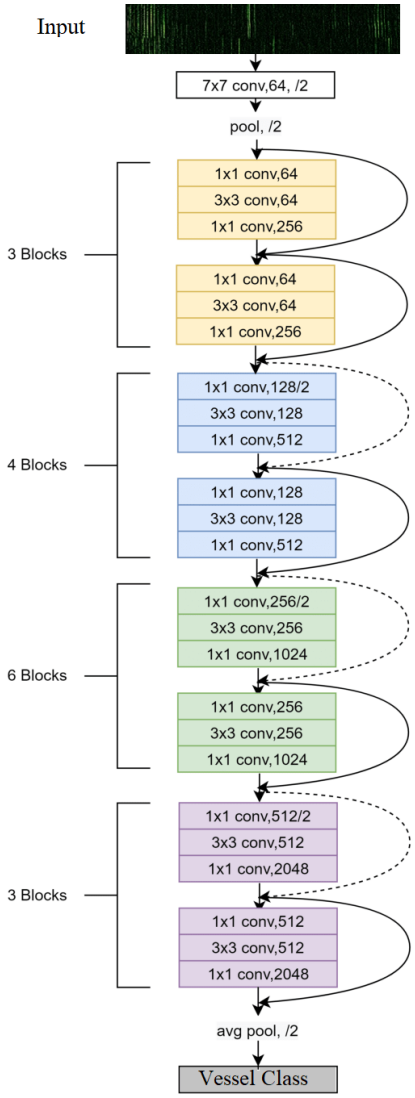

Figure 3. Modified ResNet50 architecture

connected layer with 1,000 neurons was replaced with a fully connected layer containing 4 neurons, as shown in Figure 3.

Let r and c represent the number of spectrogram pixel rows and columns, respectively. The modified network accepted an $(r \times c \times 3)$ input image, where the grayscale LOFAR gram was copied across all three-color channels expected by ResNet-50. Following the removal of the original ImageNet classification layers, the ResNet-50 network produced outputs of shape $(r\alpha \times c\alpha \times f\alpha)$, where $f\alpha$ represents the number of extracted features for each spatial location $r\alpha \times c\alpha$. Using a spatial average pooling layer, the fully convolutional layer is converted into $(r\alpha \times c\alpha \times f\alpha)$ output into an f-dimensional feature vector, which was subsequently fed to the final classification layer comprising four output neurons.

A sigmoid activation function was initially employed in the modified classification layer to allow each ship-specific neuron to independently detect the presence of its target class, accommodating early experiments that considered potential overlaps in class characteristics. However, as the final task involves mutually exclusive ship classes, the activation was replaced with a softmax function to ensure proper probability normalization and single-class prediction. Minimal preprocessing was applied to the training images: pixel intensity values in the range [0,255] were normalized. Additionally, data augmentation was performed by resizing the LOFAR grams from (50 × 4000 × 1) to (224 × 224 × 3), where 50 and 4000 are the time instances and frequency bins, respectively.

The network was optimized using stochastic gradient descent with momentum, starting with a learning rate of 0.001. Training proceeded for ten epochs with a 64-mini-batch size, leveraging GPU acceleration. The experiments were conducted on a system equipped with an AMD Ryzen 5 3600 six-core CPU, 32 GB of RAM, a 500 GB SSD, and an NVIDIA GTX 1660 SUPER GPU with 6 GB of RAM.

3.3 Data Preprocessing

To generate LOFAR grams, the publicly available DeepShip dataset [18] was systematically processed. The workflow begins with loading the audio recordings and setting the main CAPSE parameters. Each segment used a 16,000-sample window with 50% overlap, and the sampling frequency was 8 kHz.

For LOFAR gram construction, the acoustic signals were divided into segments, and a Hanning window was applied to lower spectral leakage. Another real FFT was calculated for every segment, and the energy at each frequency bin was normalized. The first half of the frequency bins were initially stored, followed by additional FFT operations along each spectral column. Subsequently, squaring the magnitudes, the peak value of every column was retained. The final spectral representation was arranged as a 50×4000 row-vector matrix in logarithmic scale and saved as PNG images. Figure 4 presents examples of zoomed LOFAR gram images (0–1600 Hz) for the four classes using sample data from DeepShip.

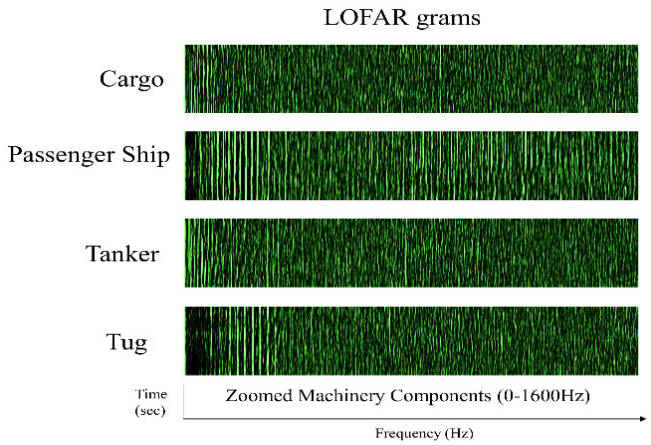

Figure 4. Sample of LOFAR grams on DeepShip dataset for Cargo, Passenger Ship, Tanker and Tug

4. RESULTS AND DISCUSSION

This segment portrays and investigates the results of the proposed approach for passive target classification based on sounds. The model's performance is evaluated on benchmark datasets, emphasizing classification accuracy and the benefits of using CAPSE-enhanced spectra with

ResNet50. Figures 5 and 6 illustrate the training accuracy and loss curves.

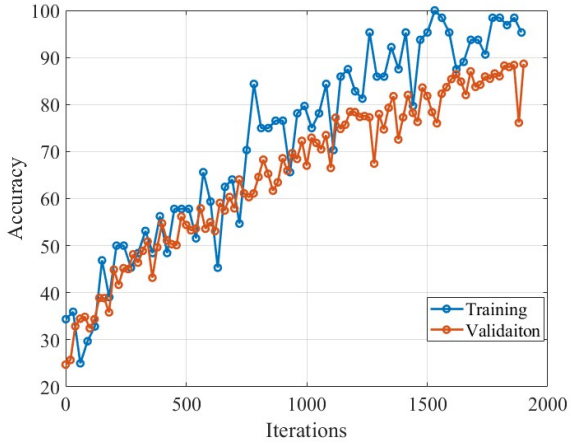

Figure 5. Accuracy curves for training and validation

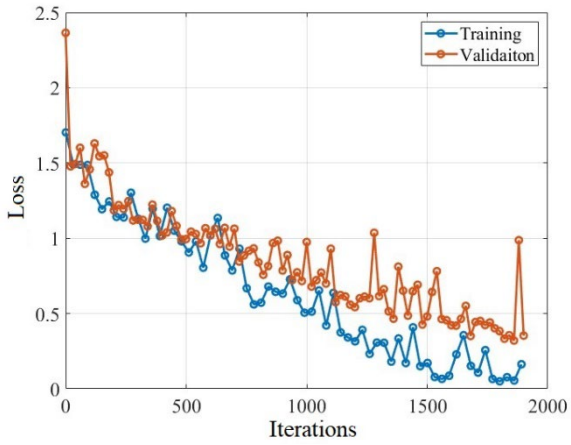

Figure 6. Loss curves for training and validation

4.1 Classification performance

Table 1 summarizes the dataset used to evaluate classification performance, listing the total number of test samples for each ship class. Regardless of differences in sample sizes, the model generalizes well among every class, maintaining effective performance even for underrepresented classes like the Tug, as illustrated in the confusion matrix in Figure 7.

Table 1. Description of DeepShip dataset into classes [18]

Class Description	Total samples
Cargo	4242
PassengerShip	4643
Tanker	4452
Tug	4054

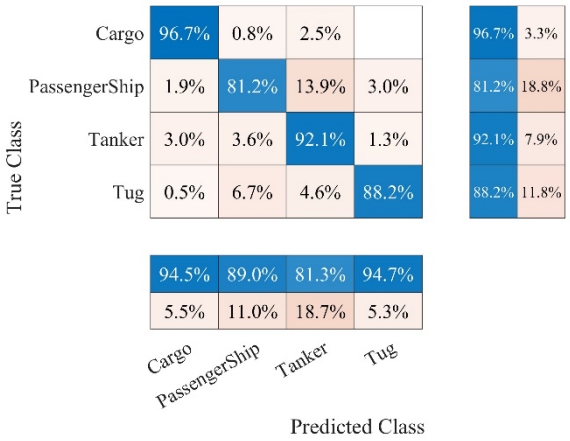

Figure 7. Confusion matrix for DeepShip dataset

Table 2 summarizes the classification results, showing that the model performs strongly across all ship types. The highest accuracy is achieved for the Cargo class at 96.70%, followed by the Tanker class at 92.07%. The Passenger and Tug classes exhibit slightly lower accuracies of 81.18% and 88.16%, respectively, indicating consistent performance regardless of class sample size.

Table 2. Classification performance of ViT network

Label	Accuracy	Precision	Recall	F1-Score
Cargo	96.70	94.75	96.70	95.72
Passenger Ship	81.18	87.95	81.18	84.43
Tanker	92.07	81.39	92.07	86.40
Tug	88.16	95.28	88.16	91.58
Average	89.53	89.84	89.53	89.53

Additional metrics—precision, recall, and F1-score further confirm the method's strength. F1-scores, which balance precision and recall, range from 84.43% for Passenger ships to 95.72% for Cargo, reflecting reliable classification across varied acoustic profiles. The average F1-score of 89.53% demonstrates strong generalization to unseen test data, underscoring the model's suitability for passive acoustic classification. Minor performance variations, particularly for Passenger ships due to lower acoustic emissions, do not detract from the overall strong classification capability across all ship types.

4.2 Features Visualization

T-SNE algorithm was used to visualize the model's feature extraction. Multi-dimensional features of ships noise were mapped into two dimensions to assess class separability. Figure 8 shows the t-SNE plot for the untrained network, where samples are dispersed with no discernible structure. After training, Figure 9 illustrates a clear division of groups, with most test samples correctly grouped according to their labels. A few misclassified points likely reflect weak or ambiguous target signatures, highlighting areas for potential improvement. Overall, this visualization demonstrates that the model effectively captures discriminative features, producing well-defined class clusters in feature space.

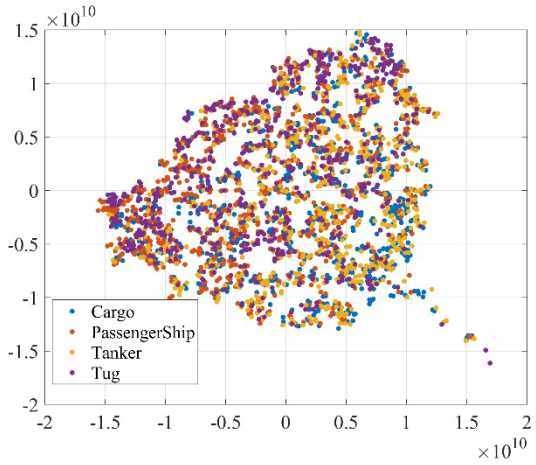

Figure 8. tSNE plot of untrained network

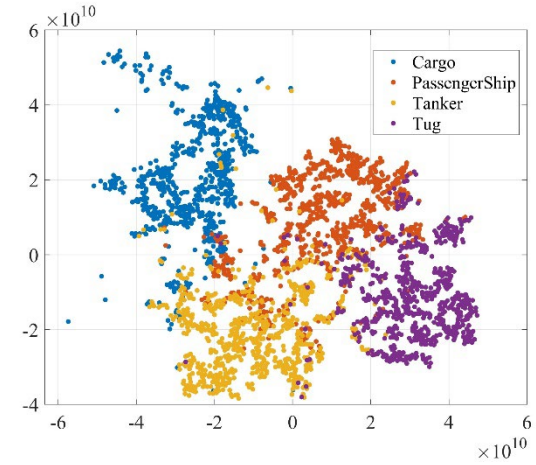

Figure 9. tSNE plot of trained network

4.3 Evaluation against state-of-art models

Table 3 gives an assessment of the proposed model and various state-of-art methods reported in the literature, including DRACNN, AGNet, SNA-Net, and SCBAN. The proposed model attains a better classification performance with an accuracy of **89.93%**, over the listed baseline models. Notably, it surpasses DRACNN [38], a similar model, with a marginal improvement, while offering a significant performance gain over AGNet [10], SNA-Net [39], and SCBAN [18] by more than 11%, 10%, and 12%, respectively. This demonstrates the effectiveness of the proposed CAPSE preprocessing and ResNet50 classification pipeline under varying acoustic and environmental conditions.

Table 3. Comparison table with research works from literature

Models	**Accuracy (%)**
DRACNN [38]	89.20%
AGNet [10]	77.09%
SNA-Net [40]	78.25%
SCBAN [18]	77.53%
Proposed Model	**89.93%**

5. CONCLUSIONS AND FUTURE WORK

A deep learning framework combining CAPSE for signal processing and ResNet50 as the classifier is proposed to improve the identification of radiated noise of marine ships. The model achieves improved performance, with an accuracy and F1-score of 89.53%, comparable to state-of-art methods. It effectively extracts discriminative features while maintaining low computational complexity, making it suitable for resource-constrained and real-time applications. These results demonstrate the approach's ability to balance classification accuracy with efficiency.

Publicly available datasets are limited, and the recorded ship types are insufficient. It is difficult to fully assess the strength of the model under numerous environmental conditions. To attend to the limitations posed by scarce real-world underwater acoustic datasets, future work will focus on synthetic dataset generation through mathematical modeling of ship noise, ambient noise, and the propagation channel. This approach allows the creation of diverse and realistic acoustic scenarios, including unique and extreme conditions, thereby enhancing model generalization. Synthetic data also enables controlled experimentation, balanced class representation, and cost-effective scalability. In parallel, domain adaptation techniques such as Demodulation of Enveloped Noise (DEMON) and intrinsic transient noise characteristics associated with ships will be explored to further improve the robustness and adaptability of classification models to unseen operational and environmental conditions.

REFERENCES

[1] L. C. F. F. Domingos, P. E. Santos, P. S. M. M. Skelton, R. S. A. A. Brinkworth, and K. Sammut, "A Survey of Underwater Acoustic Data Classification Methods Using Deep Learning for Shoreline Surveillance," *Sensors*, vol. 22, no. 6, p. 2181, Mar. 2022, doi: 10.3390/s22062181.

[2] M. Thomas, B. Martin, K. Kowarski, B. Gaudet, and S. Matwin, "Marine mammal species classification using convolutional neural networks and a novel acoustic representation," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2019, Würzburg, Germany, September 16--20, 2019, Proceedings, Part III*, 2020, pp. 290–305.

[3] L. Bjørnø, "Underwater acoustic measurements and their applications," in *Applied underwater acoustics*, Elsevier, 2017, pp. 889–947.

[4] M. F. McKenna *et al.*, "Understanding vessel noise across a network of marine protected areas," *Environ. Monit. Assess.*, vol. 196, no. 4, 2024, doi: 10.1007/s10661-024-12497-2.

[5] P. Cnn, "Marine Radar Small Target Classification Based on Block-Whitened Time – Frequency Spectrogram," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–11, 2023, doi: 10.1109/TGRS.2023.3240693.

[6] S. Siddagangaiah *et al.*, "A complexity-based approach for the detection of weak signals in ocean ambient noise," *Entropy*, vol. 18, no. 3, p. 101, 2016.

[7] M. A. Aslam *et al.*, "Underwater sound classification

using learning based methods: A review," *Expert Syst. Appl.*, vol. 255, no. June, 2024, doi: 10.1016/j.eswa.2024.124498.

[8] X. Luo, L. Chen, H. Zhou, and H. Cao, "A Survey of Underwater Acoustic Target Recognition Methods Based on Machine Learning," *J. Mar. Sci. Eng.*, vol. 11, no. 2, 2023, doi: 10.3390/jmse11020384.

[9] H. R. Gupta, "Power Spectrum Estimation using Welch Method for various Window Techniques," vol. 2, no. 6, pp. 389–392, 2013.

[10] Y. Xie, J. Ren, and J. Xu, "Adaptive ship-radiated noise recognition with learnable fine-grained wavelet transform," *Ocean Eng.*, vol. 265, p. 112626, 2022.

[11] D. Liu, H. Yang, W. Hou, and B. Wang, "A Novel Underwater Acoustic Target Recognition Method Based on MFCC and RACNN," *Sensors*, vol. 24, no. 1, p. 273, 2024.

[12] N. Müller, J. Reermann, and T. Meisen, "Navigating the Depths: A Comprehensive Survey of Deep Learning for Passive Underwater Acoustic Target Recognition," *IEEE Access*, 2024.

[13] Y. Zeng, M. Zhang, F. Han, Y. Gong, and J. Zhang, "Spectrum Analysis and Convolutional Neural Network for Automatic Modulation Recognition," *IEEE Wirel. Commun. Lett.*, vol. 8, no. 3, pp. 929–932, 2019, doi: 10.1109/LWC.2019.2900247.

[14] J. de C. V. Fernandes, N. N. de Moura Junior, and J. M. de Seixas, "Deep Learning Models for Passive Sonar Signal Classification of Military Data," *Remote Sens.*, vol. 14, no. 11, 2022, doi: 10.3390/rs14112648.

[15] X. Lin, R. Dong, Y. Zhao, and R. Wang, "Efficient ship noise classification with positive incentive noise and fused features using a simple convolutional network," *Sci. Rep.*, vol. 13, no. 1, pp. 1–13, 2023, doi: 10.1038/s41598-023-45245-6.

[16] T. A. Lampert and S. E. M. O'Keefe, "On the detection of tracks in spectrogram images," *Pattern Recognit.*, vol. 46, no. 5, pp. 1396–1408, 2013, doi: 10.1016/j.patcog.2012.11.009.

[17] H. Lan, P. R. White, N. Li, J. Li, and D. Sun, "Coherently averaged power spectral estimate for signal detection," *Signal Processing*, vol. 169, p. 107414, 2020.

[18] 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 Syst. Appl.*, vol. 183, p. 115270, 2021.

[19] L. C. F. Domingos, P. E. Santos, P. S. M. Skelton, R. S. A. Brinkworth, and K. Sammut, "A Survey of Underwater Acoustic Data Classification Methods Using Deep Learning for Shoreline Surveillance," pp. 1–30, 2022.

[20] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, Jun. 1967, doi: 10.1109/TAU.1967.1161901.

[21] P. M. David and B. Chapron, "Underwater acoustic signal analysis with wavelet process," *J. Acoust. Soc. Am.*, vol. 87, no. 5, pp. 2118–2121, 1990.

[22] T. Lim, K. Bae, C. Hwang, and H. Lee, "Classification of underwater transient signals using MFCC feature vector," *2007 9th Int. Symp. Signal Process. its Appl. ISSPA 2007, Proc.*, pp. 8–11, 2007, doi: 10.1109/ISSPA.2007.4555521.

[23] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling," *Int. Symp. Music Inf. Retr.*, vol. 28, p. 11p., 2000, doi: 10.1.1.11.9216.

[24] P. Singh, G. Saha, and M. Sahidullah, "Non-linear frequency warping using constant-Q transformation for speech emotion recognition," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, pp. 1–6.

[25] J. Park and D.-J. Jung, "Deep Convolutional Neural Network Architectures for Tonal Frequency Identification in a Lofargram," *Int. J. Control. Autom. Syst.*, vol. 19, no. 2, pp. 1103–1112, Feb. 2021, doi: 10.1007/s12555-019-1014-4.

[26] F. Yuan, X. Ke, and E. Cheng, "Joint representation and recognition for ship-radiated noise based on multimodal deep learning," *J. Mar. Sci. Eng.*, vol. 7, no. 11, p. 380, 2019.

[27] X. Luo, M. Zhang, T. Liu, M. Huang, and X. Xu, "An underwater acoustic target recognition method based on spectrograms with different resolutions," *J. Mar. Sci. Eng.*, vol. 9, no. 11, p. 1246, 2021.

[28] G. Song, X. Guo, W. Wang, Q. Ren, J. Li, and L. Ma, "A machine learning-based underwater noise classification method," *Appl. Acoust.*, vol. 184, p. 108333, 2021.

[29] N. N. De Moura and J. M. De Seixas, "Novelty detection in passive SONAR systems using support vector machines," *2015 Latin-America Congr. Comput. Intell. LA-CCI 2015*, 2016, doi: 10.1109/LA-CCI.2015.7435957.

[30] B. M. Sherin and M. H. Supriya, "Selection and parameter optimization of SVM kernel function for underwater target classification," *2015 IEEE Underw. Technol. UT 2015*, pp. 1–5, 2015, doi: 10.1109/UT.2015.7108260.

[31] A. E. Hegazy, M. A. Makhlouf, and G. S. El-Tawel, "Improved salp swarm algorithm for feature selection," *J. King Saud Univ. Inf. Sci.*, vol. 32, no. 3, pp. 335–344, 2020.

[32] G. Hu, K. Wang, and L. Liu, "Underwater acoustic target recognition based on depthwise separable convolution neural networks," *Sensors*, vol. 21, no. 4, pp. 1–20, 2021, doi: 10.3390/s21041429.

[33] X. Cao, R. Togneri, X. Zhang, and Y. Yu, "Convolutional neural network with second-order pooling for underwater target classification," *IEEE Sens. J.*, vol. 19, no. 8, pp. 3058–3066, 2018.

[34] L. Cinelli, G. Chaves, and M. Lima, "Vessel classification through convolutional neural networks using passive sonar spectrogram images," *Proc. Simpósio Bras. Telecomunicações e Process. Sinais (SBrT 2018), Armação Buzios, Brazil*, pp. 21–25, 2018.

[35] P. A. A. Esquef, L. W. P. Biscainho, and V. Välimäki, "An efficient algorithm for the restoration of audio signals corrupted with low-frequency pulses," *J. Audio Eng. Soc.*, vol. 51, no. 6, pp. 502–517, 2003.

[36] S. Feng, K. Jiang, and X. Kong, "A line spectrum detector based on improved coherent power spectrum estimation," *J. Phys. Conf. Ser.*, vol. 1971, no. 1,

2021, doi: 10.1088/1742-6596/1971/1/012006.

[37] S. Mascarenhas and M. Agarwal, "A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification," in *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, IEEE, 2021, pp. 96–99.

[38] F. Ji, J. Ni, G. Li, L. Liu, and Y. Wang, "Underwater acoustic target recognition based on deep residual attention convolutional neural network," *J. Mar. Sci. Eng.*, vol. 11, no. 8, p. 1626, 2023.

[39] P. Zhu, Y. Zhang, Y. Huang, C. Zhao, K. Zhao, and F. Zhou, "Underwater acoustic target recognition based on spectrum component analysis of ship radiated noise," *Appl. Acoust.*, vol. 211, no. July, p. 109552, 2023, doi: 10.1016/j.apacoust.2023.109552.

[40] P. Zhu, Y. Zhang, Y. Huang, C. Zhao, K. Zhao, and F. Zhou, "Underwater acoustic target recognition based on spectrum component analysis of ship radiated noise," *Appl. Acoust.*, vol. 211, p. 109552, 2023.