IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 June 2025, accepted 25 June 2025, date of publication 2 July 2025, date of current version 16 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3585467

RESEARCH ARTICLE

IARA: An Underwater Acoustic Database

FÁBIO OLIVEIRA BAPTISTA DA SILVA[1,3], JÚLIO DE CASTRO VARGAS FERNANDES[2], WILLIAM SOARES FILHO[3], JOÃO BAPTISTA DE OLIVEIRA E SOUZA FILHO[1], ÂNGELA SPENGLER[4], AND NATANAEL NUNES DE MOURA JÚNIOR[1]

[1]Signal Processing Laboratory, Federal University of Rio de Janeiro, Rio de Janeiro 21941-598, Brazil
[2]National Laboratory for Scientific Computing, Petrópolis 25651-075, Brazil
[3]Underwater Acoustic Systems Group, Brazilian Navy Research Institute, Rio de Janeiro 21931-095, Brazil
[4]Petrobras: Brazilian O&G Company, Rio de Janeiro 20231-030, Brazil

Corresponding author: Fábio Oliveira Baptista da Silva (fabio.oliveira@lps.ufrj.br)

This work was supported by CAPES Transformative Agreement.

ABSTRACT Recent advancements in artificial intelligence have been driven by the availability of diverse and well-structured public databases; however public databases for passive sonar classification remain scarce. To address this gap, this paper presents a new public underwater acoustic labeled database collected during the Santos Basin Soundscape Monitoring Project carried out by PETROBRAS in compliance with conditions required by the federal environmental licensing process conducted by IBAMA. This database expands public resources, containing 129 h 34 min of audio across 1825 recordings (approximately 50 GB). In contrast, existing public databases include ShipsEar (3 h 9 min, 90 recordings) and DeepShip (47 h 4 min, 613 recordings). The dataset provides several hours of general-purpose recordings; however, in this work, only classification was explored, as this is a typical use of such databases in the literature. Baselines were established using Random Forest, MLP, and CNN classification models, with the MLP achieving the best performance: a balanced accuracy of 67.48 $\pm$ 1.24% and an F1-score of 66.72 $\pm$ 1.17%, evaluated using 5×2 cross-validation. The primary contribution of this work is the introduction of the database and its associated baselines for underwater acoustic target recognition tasks, with the aim of providing a valuable resource for the research community. The study also emphasizes the importance of proper data selection, including the use of exclusion regions to enhance signal confidence. Generalization experiments highlight the critical role of dataset diversity, demonstrating that models often struggle to generalize to other datasets. These findings underscore the necessity of diverse, well-structured databases to improve model robustness across varied acoustics scenarios. The IARA an Acoustical Recordings Archive (IARA) is available at doi.org/10.5281/zenodo.15777429 and the code used in this paper is available at github.com/labsonar/IARA

INDEX TERMS IARA, underwater acoustic dataset, ship noise dataset, underwater acoustic target recognition.

I. INTRODUCTION

The Underwater Acoustic Target Recognition (UATR) task plays a crucial role in a wide range of underwater tasks, including the inspection of offshore structures, coastal surveillance, marine life identification, and vessel detection, supporting both ecological and security studies [1].

Passive Sonar vessel classification, i.e., the task of identifying vessels based solely on the acoustic signatures of ship noise propagating through water, presents inherent challenges due to the variability in the propagation properties of the medium (such as density, compressibility, temperature, and salinity); boundary effects (such as surface/bottom interactions, multipath and obstacles); presence of numerous noise sources (such as environmental noise and vessels irrelevant to the application), and the nonlinearities and inhomogeneities encountered in the propagating medium, thereby resulting in a challenging task [2]. These noise sources include natural sounds from sea-surface agitation and weather conditions, as well as anthropogenic noises primarily related to commercial shipping and oil exploration [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Xuebo Zhang.

Due to the complexity of the underwater acoustic channel, accurately determining its exact transfer function is a challenging task. As a result, various models have been developed to estimate it, such as parabolic equation (PE), ray, normal mode (NM), and wavenumber integration (WI) models. These models, their strengths, and their limitations are discussed in [4]. However, they often require additional information that is typically not available alongside the acoustic data, such as sound speed profiles, acoustic properties of the seabed, and local water depth [4]. Moreover, these modeling techniques typically ignore various inhomogeneities, including undulating seabeds, variations in the acoustic wave properties of the bottom across the propagation path, the presence of seamounts and bubbles, and internal waves within the water column [5].

Recently, there has been an increase in the use of deep learning methods to address these complex tasks, driven primarily by the capabilities of these models to handle the highly non-linear data typical of underwater acoustics [1]. However, the effectiveness of applying these models depends heavily on the availability of extensive and diverse dataset.

Despite the critical need for data to train complex models, there is a notable scarcity of public data in the passive sonar domain [6], [7], [8], [9], [10], [11]. The limited availability of public data can be attributed to several factors, including the confidential nature of many databases and the high costs associated with underwater data acquisition [12]. Among the few available labeled datasets are (i) ShipsEar [13], which comprises 90 recordings from 11 vessel types, covering approximately 3 h of acoustical data; (ii) DeepShip [14], featuring 613 recordings spanning around 47 h and involving 265 ships across 4 different vessel type; and (iii) QiandaoEar22 [15], the first dataset for multi-target ship classification that encompasses 9 hours and 28 minutes of real-world ship-radiated noise data and, 21 hours and 58 minutes of background noise data.

This paper aims to enhance the resources available for machine learning applications in UATR, by providing a diverse set of underwater acoustic data, including ship noises and background sounds, through the release of the **I**ARA an **A**coustical **R**ecordings **A**rchive (IARA). The name IARA is an acronym inspired by a character from Tupi-Guarani mythology [16] and available at doi.org/10.5281/zenodo.15777429.

The data on IARA were collected in the context of Santos Basin Soundscape Monitoring Project (PMPAS-BS), carried out by PETROBRAS in compliance with conditions required by the federal environmental licensing process conducted by IBAMA.

The main contributions of this work are summarized as follows:

1) Open Data Archive Publication: This work presents a comprehensive acoustic data selection process that incorporates situational awareness, as well as acoustic and spectral verification, to enhance the reliability of both the data and accompanying metadata.
2) Baseline Establishment: The study establishes baselines for IARA database using Random Forest (RF), MultiLayer Perceptron (MLP), and Convolutional Neural Network (CNN) classifiers, including codes for data loading, preprocessing, and model training.
3) Classifier Performance Evaluation: The work investigates the performance of classifiers with regards to:
    a) Closest Point of Approach (CPA) Proximity: Assesses the influence of ship proximity on classification performance by evaluating data partitions that contain CPA events.
    b) CPA Presence: Evaluates the impact of CPA occurrences on classification tasks by comparing recordings with and without CPA.
    c) Geographic Impact: Examines the effect of recording location on classification tasks by combining IARA with ShipsEar recordings, which were collected near the Port of Vigo, in Spain.
    d) Depth Impact: Analyzes the influence of recording depth on classification performance by using IARA data in conjunction with the DeepShip data.

The remainder of this article is organized as follows: Section II reviews the literature on passive sonar classification and introduces relevant underwater acoustic labeled databases. Section III describes the data acquisition process and the recording apparatus, ensuring data reliability. Section IV outlines the data selection process and the constraints applied to each segment. Section V details the instance aggregation used in the experiments, along with the preprocessing steps and metrics used to evaluate the results. Section VI presents the classification experiments, including grid searches and performance comparisons. Finally, Section VII summarizes the main findings, discusses their practical implications, and proposes directions for future research.

II. RELATED WORK

The following sections provide a comprehensive review of related topics. First, we present an overview of passive sonar systems and classifiers, emphasizing their importance in underwater acoustics. Next, we discuss existing public acoustic databases, focusing on their characteristics and availability. To highlight the value of such databases, we also review recent studies that have leveraged these resources, showcasing their contributions to advancements in UATR and illustrating the potential applications of these data.

A. PASSIVE SONAR CLASSIFIERS

Passive sonar technology plays an important role in underwater surveillance, detecting, and analyzing acoustic waves. Unlike active sonar, it operates covertly, making it indispensable in military and civilian applications [17]. These systems are crucial in naval operations for submarine detection, tracking, and threat assessment while maintaining stealth [18]. Additionally, passive sonar aids in marine

research, allowing the study of marine life and monitoring environmental changes with minimal disturbance to ecosystems [19]. Ongoing advancements promise increased precision and capabilities in underwater surveillance [2].

In machine learning applied to passive sonar classification research, the prevailing approach involves the utilization of a feature extractor in conjunction with a non-linear classifier [12]. Primary feature extraction methods encompass the frequency power spectrum, usually obtained through Fast Fourier Transform (FFT) analysis over windowed signals [20], as Mel-Spectrogram (MEL) [21], and the traditional Low-Frequency Analysis and Recording (LOFAR) [22], as well as time-frequency information extracted from wavelet features [23], and features derived directly from the audio waveform, such as zero-crossing or peak-to-peak characteristics [24]. These features are typically represented and analyzed as two-dimensional images, where one axis corresponds to time, allowing the observation of how signal characteristics evolve temporally. This time-dependent visualization facilitates the identification of transient or dynamic patterns critical for classification tasks.

B. ACOUSTIC DATABASES

Many studies in the field of underwater acoustics rely on private databases, such as [25], [26], [27], [28], [29], [30], which can significantly hinder the reproducibility of proposed techniques. For instance, [26] uses a database of 103 ship passages recorded under a European Defence Agency project, which is not publicly accessible. Similarly, [27] employs a private database compiled from proprietary recordings of underwater environments for multi-label classification of heterogeneous underwater sounds. These examples highlight the challenges of validating and extending research based on databases that are not openly shared with the scientific community.

To the best of our knowledge, there are only two publicly available labeled databases for underwater acoustics ship single-target classification: ShipsEar [13] and DeepShip [14].

In 2016, ShipsEar [13] was introduced, consisting of acoustic recordings primarily collected from the coastal waters near the port of Vigo, Spain, between the autumn of 2012 and the summer of 2013. The port of Vigo is one of the largest fishing ports worldwide and experiences significant traffic in both goods and passengers.

The collected acoustic data includes two main types: background noise and ship noise. Background noise recordings capture acoustic data from the environment, while the associated metadata includes information such as temperature, humidity, wind speed, and rainfall, obtained from nearby meteorological stations. Ship noise data includes sounds from various types of vessels, such as fishing boats, trawlers, mussel boats, pilot ships, tugboats, dredgers, roll-on/roll-off ships (ro-ro), ocean liners, passenger ferries, sailboats, and motorboats, each possibly operating under different machinery conditions.

The recordings were manually labeled, with durations ranging from 9 s to 400 s. They were captured at distances up to 350 m from the hydrophone in areas where the local depth did not exceed 45 m.

The ShipsEar is widely employed in various studies, for instance, it has been used in the development of techniques involving convolutional recurrent neural networks (CRNNs) combined with 3-D Mel-spectrograms for underwater target recognition [31]. It has also been applied in the deployment of ambient noise-free Generative Adversarial Networks (AN-GANs) for denoising underwater ship engine audio signals [32], as well as in supporting deep learning-based feature extraction efforts [33].

Released in 2021, the DeepShip [14] comprises acoustic recordings captured from May 2016 to October 2018 along busy shipping routes near Vancouver, Canada's busiest port. These recordings were collected in the Strait of Georgia Delta node, an area known for its vibrant marine environment, influenced by river discharge and strong tidal currents.

The recordings in DeepShip were selected based on Automatic Identification System (AIS) data [34], with each recording representing a single vessel within a 2 km radius of the hydrophone. The recorded durations range from 6 s to 1530 s, and acquisition depths range from 141 m to 147 m. DeepShip includes classifications for four types of vessels: Tugs, Cargo, Tankers, and Passenger, offering a larger database than ShipsEar, though for vessels in deeper waters.

Recent research using DeepShip has showcased innovative methodologies in underwater acoustics. For example, the implementation of the Variable-Step Multiscale Katz Fractal Dimension (VSMKFD) introduces a novel nonlinear dynamic metric for analyzing ship-radiated noise [35]. Additionally, these data have been used to demonstrate the effectiveness of advanced machine learning algorithms for feature classification, focusing on categorizing ship types based on acoustic signatures [33].

Furthermore, both ShipsEar and DeepShip have been combined in studies, such as the introduction of the Underwater Acoustic Classification System with a Learnable Front-end (UALF) [36] and the development of a Transformer-based Deep Learning Network for Underwater Acoustic Target Recognition [37].

The studies highlighted here underscore the critical role of public databases in advancing the field of passive sonar. These works emphasize the importance of accessible data in driving innovation, supporting scientific progress, and ensuring reproducibility in research efforts.

ShipsEar, DeepShip, and IARA offer complementary strengths for underwater acoustic research. ShipsEar provides manually labeled data with diverse vessel states but is limited to shallow waters and a small number of recordings. DeepShip offers a larger dataset in deep waters with exclusion

zones to reduce contamination, though it lacks ambient noise recordings and has fewer vessel types. IARA, introduced in this work, includes a broader volume of data across shallow and deep waters, uses both fixed and mobile platforms, records ambient and vessel noise, and applies dual-zone filtering and acoustic verification to improve confidence. The use of multiple open datasets is ideal for developing and validating robust methods and reproducible research.

III. DATA ACQUISITION

To ensure clarity and avoid ambiguity, specific terminologies are defined and consistently used throughout this document from this point forward. The term Data Archive refers to a set of Data Collections (DCs), where each DC comprises data recorded under similar conditions and subjected to the same processing and selection procedures. In contrast, the term dataset refers to a specific subset of data extracted from one or more DCs, designed for use in machine learning training.

During the PMPAS-BS, various activities were conducted to allow the acquisition of acoustic data, including mobile and fixed monitoring in coastal and oceanic areas of the Santos Basin, such as:

- Mobile Oceanic Monitoring: This was conducted using autonomous navigation equipment such as gliders and freely drifting acoustic profilers.
- Coastal Fixed Monitoring: This involved deploying a shallow-water Underwater Observatory (UO) in coastal regions of the Santos Basin.
- Oceanic Fixed Monitoring: This entailed the installation of instrumented mooring lines near production units, navigation routes, and areas with lower intensity of Exploration and Production (E&P) activities.

These diverse monitoring approaches provided a large amount of data covering different acoustic environments within the Santos Basin. However, this study is focused solely on the data gathered from gliders and UO. This decision is driven by the limitations of the others recording devices: freely drifting acoustic profilers only provide spectral averages rather than time-series data, while instrumented mooring lines have short recording durations, making it challenging to establish confidence in vessel presence.

A. GLIDER

Gliders, a type of Autonomous Underwater Vehicle (AUV), are equipped with Passive Acoustic Monitoring (PAM) systems, Conductivity Temperature and Depth (CTD) sensors, and Global Positioning System (GPS) devices. These vehicles traverse predetermined paths across the oceanic regions of the Santos Basin and are remotely controlled via satellite. In the context of the PMPAS-BS, gliders are configured to record acoustic data during the descent phase of their dives, while CTD data is collected during ascent. These dives can reach depths of up to 1000 m and typically last approximately 3 h. After each dive, the gliders surface to recalibrate their navigation using GPS and to receive new instruction via satellite. Acoustic signals are recorded at a sample rate of 125 kHz with a 16-bit resolution.

B. UNDERWATER OBSERVATORY

UO are systems installed along the seabed in the coastal regions of the Santos Basin, specifically designed to record acoustic signals with the primary objective of capturing vessel noises. These observatories were strategically positioned near maritime traffic routes to maximize vessel detection and encompass a wide range of acoustic environments.

Autonomous recorders are programmed to record the acoustic signal in these observatories continuously. These recorders are equipped with hydrophones, preamplifiers, and signal processors that receive and locally store the environmental acoustical noise.

This equipment provides continuous recordings for periods in the range of 45 days, with a sampling rate of 48 kHz and 16-bit resolution. The system's frequency response is flat up to 20 kHz (± 2 dB), while the hydrophones used throughout the project exhibit sensitivities ranging from -200 dB re 1 V/µPa to -150 dB re 1 V/µPa. The local depth at these installation points ranges from 19 m to 28 m.

IV. DATA SELECTION

This section provides an overview of the data selection process used for each data DC within IARA. It begins by comparing the provided data with publicly available databases. The criteria used to define each DCs within the IARA are then explained, ensuring a structured and meaningful organization of the DC. Finally, the section presents a summary of the selection process for ship noise recordings and background noise recordings, detailing the methods and considerations applied in curating these subsets. The completed and detailed process is available in the open report IARA Selection Report (IARA-SR), accessible as supplementary material.

This data selection process relied on two key sets of information, previously computed at PMPAS-BS: situational awareness based on recorder installation points and vessel locations derived from AIS data; and spectral analysis involving 1/3-octave spectra of the median Sound Pressure Level (SPL) processed for each minute recorded, measured in dB re 1 µPa^2.

IARA comprises 129 h and 34 min of audio data across 1825 recordings, with individual lengths ranging from 2 min to 5 min. Given the size of the IARA some strategies may be employed to mitigate training time for deep learning models but 50 GB is not, nowadays, a huge amount. Table 1 presents a comparison of the number of recordings and the total recording time between publicly available labeled audio databases and IARA.

The IARA recordings are divided into eight DCs, Table 2 presents the splitting of IARA recordings per DC and by the presence of ship or background noise.

TABLE 1. Public recordings summary.

	Number of recordings	Recording Time
IARA	1825	129 h 34 min
ShipsEar	90	3 h 9 min
DeepShip	613	47 h 4 min

TABLE 2. Summary of IARA recordings.

	Number of recordings	Recording Time
A	456	34 h 17 min
B	199	15 h 26 min
C	341	26 h 45 min
D	449	34 h 33 min
E	261	10 h 52 min
F	35	2 h 55 min
G	37	2 h 45 min
H	47	1 h 57 min
Ship noise	1517	116 h 44 min
Background noise	308	12 h 50 min

The selection criteria used to each DC of IARA can be described as in Table 3

TABLE 3. Selection criteria for each DC of IARA.

DC	Acquisition	Noise	CPA	Minimal distance
A	UO	ship	yes	<250 m
B	UO	ship	no	<250 m
C	UO	ship	yes	>250 m and <500 m
D	UO	ship	no	>250 m and <500 m
E	UO	background	no	—
F	glider	ship	yes	—
G	glider	ship	no	—
H	glider	background	no	—

This section presents the selection process used in each DC of IARA splitting the selection process for recording with ship noise and recording of background noise. The steps outlined here are discussed in detail in IARA-SR, providing insights into the reasoning behind the adopted values and a comprehensive description for the techniques used.

A. SHIP NOISE

In the ship noise selection process, the primary objective is to identify recording time windows where it is highly likely that only one identifiable ship is present in the acquired data. To achieve this, the recorded data undergoes a series of steps.

The first step utilizes AIS information to identify time windows when a single ship is near the recording location, with no other vessels reported. This approach is similar to the methodology used in DeepShip, which restricts recordings to instances where a single vessel is identified in the AIS database within a 2 km radius. To further ensure that the noise in the recording originates mostly from the identified ship, this work defines two distinct regions: an inclusion region and an exclusion region.

These regions were established to ensure a clear attenuation difference between signals originating within the inclusion region and those coming from outside the exclusion region, considering the propagation loss from the source to the sensor.

The dimensions of these regions were defined as a balance between the desired attenuation difference and the availability of viable recording windows. For UO data (DCs A to D), the exclusion zone was set at 2 km, consistent with the DeepShip methodology. The inclusion region, however, varied: it was set to within 250 m for DCs A and B, and expanded to between 250 m to 500 m for DCs C and D.

For glider data (DCs F and G), recorded in quieter and deeper waters, the exclusion region was significantly larger, at 30 km, while the inclusion region was set at 10 km, reflecting a similar balance between attenuation difference and recording availability.

If a recording does not contain the CPA, it indicates that the exclusion radius was reached either at the beginning or end of the recording time window. This approach stems from the fact that recordings were only fragmented to exclude the CPA when one of the defined constraints was at risk of being violated. On the other hand, for recordings containing the CPA, the exclusion range may be underestimated. While this does not guarantee a better signal-to-noise ratio (signal-noise ratio (SNR)), it is possible that recordings with CPA exhibit improved SNR characteristics.

The second step involved an acoustical verification process, which relied on spectral analysis of the preprocessed acoustic data to assess whether the ship reported in a given recording window was detectable. Specifically, this step verified whether the energy levels increased as the ship approached. For DCs containing the CPA (A, C, and F), an automatic detector was employed to assist in this process. Additionally, data from all DCs underwent manual verification to ensure the reliability.

The final step was to identify the ship based on its AIS information and validate it across multiple AIS databases. Recordings from vessels with incompatible information were discarded to reduce potential misidentifications. This verification was based on consistency across databases regarding the ship's name, Maritime Mobile Service Identity (MMSI), and International Maritime Organization (IMO) number.

In summary, the recordings in each DC of IARA that contain identifiable vessel passages are categorized by vessel type, as presented in Table 4. Note that DCs E and H are not included, as they primarily consist of background noise and are not expected to contain dominant ship acoustic signatures.

It is important to emphasize that these steps do not guarantee that the selected data represents a pure recording of the reported ship. Potential sources of contamination include small vessels operating without AIS, which may be recorded alongside the reported ship; inaccuracies or omissions in the AIS databases; noise from other ships outside the exclusion region that may not be sufficiently attenuated; moored vessels

TABLE 4. Distribution of recordings by vessel type.

	A	B	C	D	F	G	Total
Cargo	161	23	106	107	26	30	453
Fishing	9	2	8	15	0	0	34
High-Speed Craft	0	2	2	4	0	0	8
Passenger	4	3	5	6	0	0	18
Pleasure Craft	3	0	1	5	0	0	9
Special Craft	208	146	137	232	1	1	725
Tanker	46	10	50	36	8	4	154
Tug	25	13	32	44	0	2	116
Total	456	199	341	449	35	37	1517

potentially interfering with the recording; and the presence of other unexpected anthropogenic noises, among other factors.

Despite these limitations, the dataset is considered to provide an adequate representative characterization of the reported vessels.

B. BACKGROUND NOISE

To characterize background noise, the aim is to select recording time windows where no vessels were reported within the exclusion region, under varying environmental conditions, to ensure the noise was as stationary as possible. The data undergoes several steps to increase the likelihood of this.

The first step involves establishing an exclusion region where no vessel is reported in the AIS database to ensure minimum attenuation of any vessel noise. For UO (DC E), the exclusion region was set to 10 km, while for the glider (DC H) the exclusion radius was set to 30 km.

A valid recording window is identified when these constraints are satisfied for at least 10 min. From this window, a 2.5 min segment was selected based on the most consistent energy level. This selection was made by evaluating the number of frequency bins with significant variation and the overall variation in SPL.

For UO data, weather conditions such as rainfall and wind speed were collected from nearby meteorological stations. However, no similar data was available for the glider, so the 47 recordings that met the constancy constraint make up DC H.

The data from UO was selected to ensure environmental diversity, based on factors like rainfall and sea state. While the rainfall was tracked independently, the sea state was estimated using wind speed thresholds, as outlined in [38]. A summary of the recordings for UO is presented in Table 5.

TABLE 5. Recordings in DC E by rain and sea state.

	Sea state							
	1	2	3	4	5	6	7	Total
With rain	25	25	25	25	25	25	1	151
Without rain	3	25	30	29	11	12	0	110

V. PROPOSED METHOD

This section outlines the data processing steps, the classifiers utilized, the training procedures adopted, and the methods applied to evaluate these classifiers, all in pursuit of two primary objectives. The first objective is to establish a baseline for IARA within the context of the UATR task. The second is to assess performance impacts concerning the proximity and presence of CPA, as well as the impact of the geographic location and the local depth.

A. AUDIO PROCESSING

As outlined in Section II-A, the prevailing approach for UATR tasks involves the utilization of a feature extractor in conjunction with a non-linear classifier. In this study, we employ two distinct feature extraction methods: LOFAR and MEL.

The LOFAR technique computes a Short Time Fourier Transform (STFT) on a time-windowed signal, followed by a Two-Pass Split Window (TPSW) normalization [12]. This approach effectively highlights consistent frequency patterns while minimizing transient noises, making it well-suited for detecting stable acoustic signatures in underwater environments [39].

In contrast, the MEL spectrogram provides a logarithmic representation of the frequency domain by mapping frequencies to the Mel scale, which approximates the human auditory system's response. This logarithmic representation enhances the perceptual relevance of spectral features [11].

To ensure consistency across the varying databases used, i.e. IARA, ShipsEar and DeepShip, all acoustic signals were resampled to a sampling rate of 16 kHz, as done in [40]. Then the feature extraction was conducted on resampled audio with an average over every eight windows of 1024 points without overlap, leading to one spectral feature each 0.512 s. Each spectral feature was then normalized, dividing it by its ℓ_2 norm.

These parameters were chosen to ensure that the training and evaluation processes could be completed within a feasible timeframe, given the large volume of data from the three databases, the number of training iterations involved in the study, and the computational resources available for this work.

Combining each of these spectral feature, each original audio was converted to a 2D array.

Two training methodologies vary according to the model: window-based approach, where each spectral feature was used as a sample; and image-based approach, where subarrays consisting of 32 spectral features with an overlap of 50% were utilized.

Figure 1 presents a diagram with this audio processing chain to transform each audio recording into samples for the experiments in this study.

B. CLASSIFIERS

This work uses three types of classifiers: Random Forest (RF), MLP and CNN. The hyperparameters varied in the grid search experiments for the classifiers were: for the RF, the number of trees and their maximum depth; for the MLP:

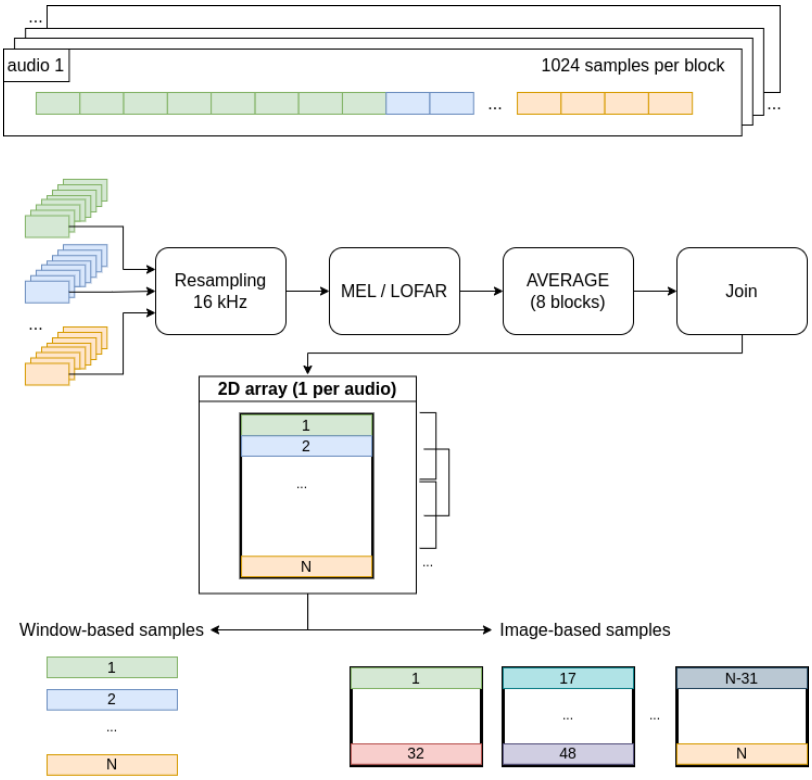


FIGURE 1. The audio signal processing chain.

the number of hidden layers, neurons per layer, activation functions for hidden and output layers, normalization layers, dropout, batch size, learning rate, and weight decay, Adam was used as the optimizer; and for the CNN: the number of neurons in each layer, activation functions in convolutional and fully connected layers, pooling type, kernel size, normalization layers, dropout, batch size, learning rate, and weight decay, also using Adam as the optimizer.

RF and MLP classifiers were trained using a window-based strategy, while CNN was trained using an image-based strategy, as detailed in Section V-A. Further details on the implementation of these classifiers are provided in the source code.

C. CLASS ASSIGNMENT

When aggregating ships for classification tasks, a hierarchical approach is often employed, consisting of two levels [34]. At the first level, ships are categorized into broad groups based on their primary function, referred to as the *general vessel type*. Examples include Cargo, Fishing, Tug, and Tanker. At the second level, these general categories are further divided into more specific subgroups, known as the *detailed vessel type*. For instance, Cargo ships can be classified into subtypes such as Bulk Carrier, Container Ship, and Ro-Ro Cargo.

In the original ShipsEar paper, classifiers were designed to distinguish groups that aggregate detailed vessel types based on their size. In contrast, the DeepShip paper focused on classifying vessels by their general vessel type.

With the IARA, direct separation by general vessel type presents significant challenges due to the high diversity of vessels. For example, the recordings include passenger vessels with lengths ranging from 10 m to 333 m, making it difficult to group them meaningfully by general type. In contrast, the vessels in the DeepShip dataset are much more homogeneous, allowing for a simpler classification approach.

Additionally, the presence of 34 different detailed vessel types in IARA complicates the classification with a strategy similar to that done in ShipsEar. In that work, detailed types were used as a proxy for vessel size; however, this assumption does not hold consistently across the IARA database, as detailed types are not always reliable indicators of size.

To address these challenges, this work takes a different approach by aggregating vessels directly based on their size, applying a ship-by-ship basis for classification. This method disregards both the general and detailed vessel types, focusing solely on vessel size to create the categories.

The exact length thresholds used in ShipsEar are not explicitly defined. Therefore, in this work, a standard size-based classification is adopted, following the classification outlined in [41]. Vessels are categorized as small if their lengths are less than 50 m, medium for vessels with lengths between 50 m and 100 m, and large for vessels exceeding

100 m. Additionally, a separate class is included to represent background noise.

The Table 6 shows the distribution of each class along each DC of IARA and the same categorization applied to the ShipsEar and the DeepShip sample data.

TABLE 6. Classes assigned in each DC of IARA.

Target	Small	Medium	Large	Background
A	113	120	223	0
B	90	73	36	0
C	104	75	162	0
D	169	129	151	0
E	0	0	0	261
F	0	1	34	0
G	0	2	35	0
H	0	0	0	47
IARA UO	476	397	572	261
IARA glider	0	3	69	47
ShipsEar	49	17	12	12
DeepShip	4	2	56	0

The DeepShip sample data and the glider data in IARA (union of DCs F, G and H) are not well represented on all targets, as shown in Table 6, making it impractical to train models under these targets. Still, it remains possible to use this data as test data for classifiers trained on ShipsEar and IARA data, thereby evaluating the accuracy of these models' predictions.

D. PERFORMANCE EVALUATION

The evaluations in this paper use the 5×2 cross-validation F test to compare the performance of two learning models [42].

The process begins with five iterations of two-fold cross-validation. In each iteration, the dataset is randomly divided into two equal-sized folds. Each algorithm is trained on one fold and tested on the other. The roles are then switched, with each algorithm being trained on the previously tested fold and tested on the previously trained fold. This process results in ten pairs of training and testing sets. The performance of the two models is compared using an F test, with 10 and 5 degrees of freedom. This test evaluates whether there is a significant difference between the algorithms by testing the null hypothesis that the distribution of the performance metrics has equal variance for both models. If the null hypothesis is rejected, it indicates a statistically significant difference in performance [42].

Each test fold was randomly split into two parts: 25% is used as an early-stop set to determine when to halt the training, while the remaining 75% serves as the final test set.

All data splits into folds were performed ensuring that samples from the same audio recording were not placed in different sets. This approach aimed to maintain the integrity and independence of the testing and training sets. For model selection, the criterion used was the highest mean Sum-Product Index (SP), defined as:

$$SP = \sqrt{\frac{1}{k}\sum_{i=1}^{k} E_i \sqrt[k]{\prod_{i=1}^{k} E_i}} \quad (1)$$

where E_i represents the detection probability for the class i and k is the number of classes in the training data. Although the Accuracy (ACC) and F1-score, both macro-averaged [43], were not used for model selection, they were calculated to provide additional insights into model performance. All metrics were estimated based on the classification performance by recording rather than by sample. Specifically, the model predicts all samples for a recording, and the most frequently predicted class is used as the predicted label for the entire recording.

VI. EXPERIMENTS

In this section, two types of experiments were conducted: grid searches and performance comparisons. All experiments were performed in Python, with the code available at github.com/labsonar/IARA.

The grid searches experiments involved varying the parameters, described in Section V-B, in a non-exhaustive manner. The performance comparison experiments involved training the models selected from the grid searches on two different datasets and evaluating their performance on both. This comparison evaluated how changing the dataset affects the performance of the models, thereby assessing their robustness to data variability.

A. BASELINE

Classifiers were trained on all IARA UO data to establish a baseline, with targets defined in Section V-C. The grid search results for each classifier are summarized below:

- RF
  - MEL: 50 forests with a maximum depth of 30.
  - LOFAR: 200 forests with a maximum depth of 20.
- MLP
  - MEL: Batch size of 32, learning rate of 1e-4, weight decay of 1e-3. Two hidden layers with 32 and 16 neurons, ReLU activation, batch normalization (1D), output layer with sigmoid activation, and dropout rate of 0.2 per layer.
  - LOFAR: Same configuration as MEL, but without dropout.
- CNN
  - MEL: Batch size of 32, learning rate of 1e-6, weight decay of 1e-3. Feature extractor with two convolutional layers (1024 and 128 neurons, LeakyReLU activation, kernel size 5, padding 1), batch normalization (2D), 2×2 max pooling, dropout rate of 0.4. The fully connected layer with 128 neurons, ReLU activation, and output layer with sigmoid activation have a dropout rate of 0.4.
  - LOFAR: Similar configuration to MEL but with 512 neurons in the first convolutional layer, ReLU

activation in convolutional layers, and 4 × 2 max pooling

Table 7 presents the performances in the test set for each of these classifiers.

TABLE 7. IARA UO results.

Classifier	Feature Extractor	SP (%)	ACC (%)	F1-score (%)
RF	MEL	62.22 ± 1.86	62.64 ± 1.84	63.79 ± 1.73
	LOFAR	56.92 ± 1.69	58.87 ± 1.68	58.00 ± 1.41
MLP	MEL	63.38 ± 1.81	64.51 ± 1.75	62.89 ± 1.68
	LOFAR	66.51 ± 1.39	67.48 ± 1.24	66.72 ± 1.17
CNN	MEL	63.52 ± 2.26	64.99 ± 2.09	63.04 ± 2.02
	LOFAR	66.05 ± 1.90	67.02 ± 1.78	66.29 ± 2.13

Table 8 presents the statistical results of the 5×2 cross-validation F-test for SP across all pairwise combinations of model and feature extractor. Each cell in the table contains the p-value and the corresponding F-value (in parentheses) resulting from the comparison between the model-feature extractor combination in the respective row and the one in the respective column.

The F-test rejects the null hypothesis, i.e. that the performance metric distributions have equal variance V-D, for SP when comparing any model with the LOFAR RF. This indicates that this combination produces results with statistically significant differences. As shown in Table 7, this combination also exhibits the worst performance. Additionally, a significant difference is observed only in the comparison between LOFAR MLP and MEL RF, considering a confidence level of 0.05.

Given that the results for LOFAR did not demonstrate a significant improvement over MEL, and for RF exhibited the opposite trend, the subsequent sections will focus exclusively on MEL to optimize training time and reduce the number of models trained. All models trained on UO data will adhere to these models' configuration.

As described in Section V-D, the F test was performed for a metric such as SP, which was calculated by evaluating the classification performance of the models when classifying an entire file. Specifically, a file was assigned the most frequent label observed across all its extracted features.

As an additional evaluation, it is possible to enforce an additional constraint: a file is only labeled by the model if a minimum proportion of its segments are assigned the same label. In this approach, a file that fails to meet this classification ratio constraint is labeled as ''unknown'' and excluded from the analysis. This method allows for evaluating the performance of the model on files where it demonstrates confidence in the classification. It is expected that as this constraint becomes more stringent, the performance (SP) will improve, although the fraction of files classified (File Ratio) will decrease.

Table 9 presents the metrics for each model alongside the classification ratio constraint that achieves the highest, for details see Section SP. Additionally, the File Ratio, representing the proportion of files still classified under these constraints, is reported. The analysis reveals consistent improvements across all models. However, while the stricter constraints reduce the number of classified files to half or less, it is noteworthy that the CNN model, despite achieving a higher classification ratio, maintains a greater File Ratio. This result highlights the CNN model as a more precise classifier.

B. CPA PROXIMITY

This experiment is a performance comparison of models trained on two datasets: data from DC *A* and data from DC *C*. Both DCs contain recordings with CPA, but DC *A* includes CPA distances less than 250 m, while DC *C* includes distances ranging from 250 m to 500 m. Therefore, DC *A* expects a better SNR since the exclusion range is the same, and the ship is closer to the sensor. Table 9 shows the metrics for each classifier in each combination of training and evaluation datasets.

Table 10 shows the metrics for each classifier in each combination of training and evaluation datasets, but the f1 score was omitted because it presents the same indicators as the metrics presented.

A similar result was observed across all classifiers: when evaluated on the same DC they were trained on, models trained on DC *A* achieved significantly better results than those trained on DC *C*, indicating that the data from DC *A* may have a higher SNR.

When the model trained on DC *A* evaluated data from DC *C*, its performance degraded but still achieved results similar to those trained on DC *C*, suggesting that models trained in better SNR conditions can extrapolate to worse SNR scenarios.

Conversely, when the model trained on DC *C* evaluated data from DC *A*, the performance was statistically similar, indicating that the model could learn in worse SNR scenarios and perform adequately in better SNR conditions. However, it is important to note that this performance was significantly worse than that of the model trained on the DC *A*.

C. CPA PRESENCE

This performance comparison experiment evaluates models trained on two datasets: one using data from DCs *A* and *C* (recordings with CPA) and the other using data from DCs *B* and *D* (recordings without CPA). The key difference lies in the presence of the CPA in the recordings, which corresponds to the closest point of approach of a sound source relative to the receiver.

At the CPA, the Lloyd's mirror effect becomes particularly pronounced. The Lloyd's mirror effect is an interference phenomenon that occurs when sound waves reflect off the water surface and combine with the direct path waves [38]. This interaction results in a pattern of constructive and destructive interference, producing peaks and valleys in the received signal. At the CPA, these interference patterns are most frequent and intense due to the minimal angle between the direct and reflected paths, amplifying the effect. This creates unique acoustic signatures that may significantly

TABLE 8. 5 × 2 cv F test result, p-value (F value).

Feature Extractor	Classifier	MEL MLP	MEL CNN	LOFAR RF	LOFAR MLP	LOFAR CNN
MEL	RF	0.57 (0.94)	0.49 (1.10)	0.01 (9.90)	0.00 (25.85)	0.12 (3.03)
	MLP	-	0.74 (0.64)	0.03 (6.11)	0.19 (2.27)	0.20 (2.18)
	CNN	-	-	0.02 (7.35)	0.25 (1.89)	0.36 (1.45)
LOFAR	RF	-	-	-	0.00 (50.05)	0.00 (15.12)
	MLP	-	-	-	-	0.76 (0.62)

TABLE 9. Confidence analysis for MEL classifiers.

Classifier	Classification Ratio (%)	SP (%)	ACC (%)	F1-score (%)	File Ratio (%)
RF	88	72.29 ± 3.86	73.23 ± 3.62	75.12 ± 3.67	38.12
MLP	85	74.18 ± 4.04	76.26 ± 2.99	76.74 ± 3.42	39.36
CNN	93	70.04 ± 2.86	72.07 ± 2.36	71.18 ± 2.68	52.11

TABLE 10. CPA proximity results.

Classifier	Training set	Evaluated set A SP (%)	A ACC (%)	C SP (%)	C ACC (%)
RF	A	57.49 ± 3.10	59.58 ± 2.88	47.63 ± 5.44	51.39 ± 4.92
	C	48.02 ± 3.99	51.33 ± 2.66	46.08 ± 5.80	50.77 ± 4.55
MLP	A	67.25 ± 2.44	67.64 ± 2.39	59.08 ± 6.51	60.11 ± 5.68
	C	59.59 ± 5.22	59.96 ± 5.13	58.08 ± 3.70	59.02 ± 3.31
CNN	A	63.28 ± 1.95	63.78 ± 1.79	57.85 ± 5.36	59.13 ± 4.72
	C	53.30 ± 5.44	54.08 ± 5.01	56.23 ± 4.36	57.17 ± 4.09

impact the performance of models trained with or without such data.

It is important to note, as mentioned in Section IV-A, that recordings containing the CPA might exhibit a higher SNR. This is due to the more probable proximity of other vessels at the boundary of the exclusion area in recordings that do not include the CPA.

Table 11 shows the metrics for each classifier in each combination of training and evaluation datasets, but the f1 score was omitted because it presents the same indicators as the metrics presented.

TABLE 11. CPA presence results.

Classifier	Training set	Evaluated set with CPA SP (%)	with CPA ACC (%)	without CPA SP (%)	without CPA ACC (%)
RF	with CPA	58.98 ± 2.18	60.40 ± 2.10	47.77 ± 3.33	48.65 ± 2.88
	without CPA	56.29 ± 3.04	56.41 ± 3.08	50.69 ± 3.07	51.56 ± 2.71
MLP	with CPA	66.50 ± 2.31	66.86 ± 2.19	50.22 ± 2.31	50.61 ± 2.31
	without CPA	58.49 ± 2.94	59.08 ± 2.83	53.57 ± 4.10	53.82 ± 4.00
CNN	with CPA	64.65 ± 2.59	65.10 ± 2.55	49.76 ± 2.81	50.20 ± 2.72
	without CPA	52.94 ± 2.92	54.94 ± 2.04	49.69 ± 2.48	50.60 ± 2.51

A similar result was observed across all classifiers: when evaluated on the test set of the datasets they were trained on, models trained with CPA achieved better results than those trained without CPA. This may indicate that scenarios with CPA indeed have better SNR conditions but it may also indicate that the presence of CPA could make ship classification easier.

When models trained with CPA were evaluated on data without CPA, their performance degraded, resulting in a lower mean but still comparable to those trained without CPA. This suggests that models trained with CPA can generalize to scenarios without CPA. This finding is significant, as recordings are typically made closer to vessels, but inference in practice, especially in military applications, often needs to be done in approaching scenarios.

Conversely, when models trained without CPA were evaluated on data with CPA, the performance increase in mean but keeping overlapping with the originals give the error margins. This indicates that models can learn from scenarios without CPA and perform adequately under CPA conditions. However, it is important to note that this performance was significantly worse than that of models trained with CPA, except for the RF classifier, where there was some overlap in performance.

D. GEOGRAPHIC IMPACT

In this section, a performance comparison experiment, as described in Section VI, is conducted between models trained on the ShipsEar and IARA UO datasets. Both datasets were recorded in shallow water, with depths ranging from 4 m to 20 m for ShipsEar and 19 m to 28 m for IARA.

The ShipsEar dataset consists of recordings captured near a fishing port characterized by goods and passenger traffic, whereas the IARA dataset comprises recordings collected near maritime traffic routes, away from ports and docks. This experiment evaluates the ability of models to maintain performance when evaluated on the alternative dataset. Since the datasets were collected under similar conditions but in different locations, this analysis is referred here as the *geographic impact* experiment. However, it is important to note that differences in the environment, vessel types, recording scenarios, seabed characteristics, and others contribute additional variability to the experiment.

To ensure a fair performance comparison, a grid search experiment was performed to determine the optimal configuration for each model on each dataset. For IARA UO, this grid search is the baseline experiment detailed in Section VI-A. For the ShipsEar dataset, the results of the grid search, including the best parameters for each classifier, are summarized below:

- RF: 8 forests with a maximum depth of 5.
- MLP: Batch size of 30, learning rate of 1e-4, without weight decay. A hidden layer with 64 neurons, PReLU activation, batch normalization (1D), and an output layer with sigmoid activation, without dropout rate.
- CNN: Batch size of 32, learning rate of 1e-4, weight decay of 1e-3. Feature extractor with two convolutional layers (256 and 32 neurons, ReLU activation, kernel size 7, padding 2), batch normalization (2D), 2×2 max pooling, without dropout. Two fully connected layers with 64 and 32 neurons, PReLU activation, and an output layer with sigmoid activation, without dropout.

Table 12 presents the results for each classifier. It can be noted that the standard deviation is larger than in previous experiments, likely due to the smaller number of recordings, which can lead to more unstable or irregular folds. The mean performance is lower than the IARA data, but the results overlap within the margins of error. It is worth noting that this error margin is larger across all metrics, but especially in SP, likely due to the geometric characteristics of this metric.

TABLE 12. ShipsEar results.

Classifier	SP (%)	ACC (%)	F1-score (%)
RF	47.37 ± 17.46	53.78 ± 7.05	50.86 ± 7.97
MLP	53.85 ± 19.64	61.01 ± 9.36	61.31 ± 9.64
CNN	52.98 ± 19.18	62.22 ± 6.89	61.46 ± 6.91

Table 13 shows the results of this performance comparison experiment.

TABLE 13. Geographic variation results.

		Evaluated set			
		IARA		ShipsEar	
Classifier	Training set	SP (%)	ACC (%)	SP (%)	ACC (%)
RF	IARA	62.22 ± 1.86	62.64 ± 1.84	0.00 ± 0.00	29.69 ± 4.05
	ShipsEar	22.17 ± 11.65	32.82 ± 2.95	47.37 ± 17.46	53.78 ± 7.05
MLP	IARA	63.38 ± 1.81	64.51 ± 1.75	0.00 ± 0.00	23.08 ± 8.95
	ShipsEar	21.22 ± 11.87	33.01 ± 4.93	53.85 ± 19.64	61.01 ± 9.36
CNN	IARA	63.52 ± 2.26	64.99 ± 2.09	8.98 ± 18.18	31.23 ± 10.97
	ShipsEar	19.11 ± 13.25	35.81 ± 4.98	52.98 ± 19.18	62.22 ± 6.89

The model trained on ShipsEar data showed a significant reduction in performance, even considering the larger margin of error, with metrics dropping from approximately 54-62% ACC to 33-36% when evaluated on IARA data.

The model trained on IARA data exhibited an even more significant reduction in metrics, from around 62-65% ACC to 23-31% when tested on ShipsEar data. It even led to an SP value of 0, indicating that at least one class was completely misclassified. In fact, most ShipsEar data was classified as small ships, as shown in the confusion matrices in Figure 2.

This result strongly suggests that models trained on IARA data could not extrapolate to classify ship noise in the ShipsEar background; instead, they classified the ShipsEar background noise as small ships. It is possible that the background noise on the ShipsEar has some contamination with traffic from small vessels due to the proximity of the port.

This is an important finding as neither model was robust to geographic variation. It underscores the significance of diverse public datasets, as achieving robust model performance across different scenarios and backgrounds requires data collected in the most varied conditions possible.

E. LOCAL DEPTH

To evaluate the impact of local depth, specifically the difference between recordings from shallow and deep water, on model performance, an additional experiment is conducted. As the publicly available datasets do not encompass all the target classes considered in this study, as described in Section V-C, the models trained in the previous section, using ShipsEar and IARA UO, will be used to predict data from the DeepShip sample and IARA glider. This experiment aims to assess whether models trained in shallow water data can maintain their performance when evaluated on data collected in deep water environments.

The DeepShip dataset was collected at depths around 150 m (typically, depths above 180 m are considered deep water [38]), while the IARA recordings were made at depths exceeding 1000 m. Data from DeepShip were gathered using a stationary platform, whereas IARA utilized a glider, representing a moving platform.

Due to the significant imbalance in the data and the incomplete representation of all classes, as shown in Table 6, any macro-averaged metric would be difficult to interpret effectively. Given this, the evaluation will be conducted using the Unbalanced Accuracy (UACC), or micro-averaged accuracy, which better accounts for class imbalances [43].

Table 14 presents the performance of the models trained on the ShipsEar and IARA UO datasets when evaluated on their respective training sets, establishing a reference performance for UACC. Additionally, the table reports the performance of these models in this experiment when evaluating the DeepShip and IARA glider datasets.

TABLE 14. Local depth results.

		ShipsEar UACC (%)	IARA UO UACC (%)
Training Set	RF	54.55 ± 9.58	62.57 ± 1.67
	MLP	67.27 ± 10.57	62.49 ± 1.63
	CNN	69.50 ± 6.50	62.00 ± 2.08
DeepShip	RF	65.97 ± 18.23	78.71 ± 1.88
	MLP	14.03 ± 18.05	42.58 ± 10.23
	CNN	34.11 ± 32.29	30.36 ± 5.53
IARA glider	RF	70.59 ± 15.07	52.94 ± 3.23
	MLP	47.14 ± 10.09	71.60 ± 9.36
	CNN	62.86 ± 20.67	54.96 ± 9.31

The MLP and CNN models showed a significant decrease in performance when evaluated with DeepShip data, with larger standard deviation. On the other hand, both RF models showed an improved performance when evaluating DeepShip data. Similar results were seen when evaluating glider data for

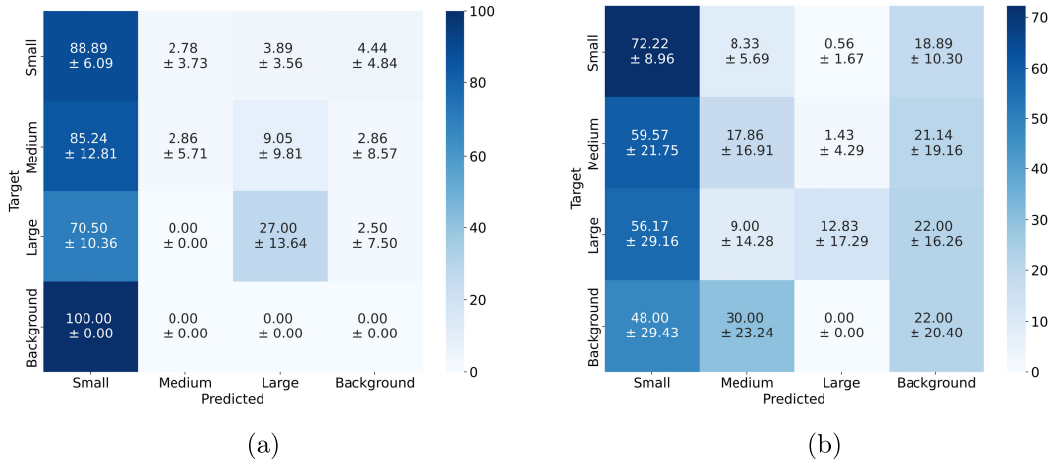


FIGURE 2. Confusion matrix for (a) RF and (b) CNN training in IARA evaluated in ShipsEar.

models trained on ShipsEar, but performance varied for UO trained models, with an increase for MLP and a decrease for RF and CNN.

The MLP and CNN models exhibited a significant decrease in performance when evaluated on the DeepShip database, accompanied by larger standard deviations. In contrast, both RF models demonstrated improved performance when evaluated on the DeepShip. Similar trends were observed when evaluating glider data using models trained on the ShipsEar dataset. However, for models trained on UO dataset, performance varied, with an increase observed for the MLP model, while a decrease was noted for the RF and CNN models.

VII. CONCLUSION

The main contribution of this work is the creation of the largest public underwater acoustical labeled dataset called IARA. IARA contains 129 h and 34 min of audio data from 1825 recordings, including ship and background noises in different environmental conditions. This provision could aid in future research in the field of underwater acoustic.

As the data of IARA are divided into DC based on different recording constraints was possible to evaluate the impact of proximity and presence of CPA and verify that models trained in data with CPA and lower CPA distances achieve significantly better results and could extrapolate to scenarios without CPA or with higher CPA distances keeping similar performances to models directly trained in these scenarios.

The study also highlights the importance of dataset diversity in achieving robust model performance across different scenarios and backgrounds by verifying that models trained in ShipsEar or IARA could not maintain their performance when evaluated over another dataset. These findings underscore the need for continued exploration of diverse public datasets to enhance the generalizability of classification models dedicated to underwater acoustic applications.

Overall, this research aims to contribute to the field of UATR and serve as a basis for future studies that use: the data selection process, the Data Archive (DA), the baselines or the methodology presented. Future works should focus on expanding public data and addressing the challenges posed by geographic and environmental variability.

REFERENCES

[1] 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," *Sensors*, vol. 22, no. 6, p. 2181, Mar. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/6/2181

[2] C. Satheesh, S. Kamal, A. Mujeeb, and M. H. Supriya, "Passive sonar target classification using deep generative β-VAE," *IEEE Signal Process. Lett.*, vol. 28, pp. 808–812, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9397277

[3] M. Mustonen, A. Klauson, T. Folégot, and D. Clorennec, "Natural sound estimation in shallow water near shipping lanes," *J. Acoust. Soc. Amer.*, vol. 147, no. 2, pp. EL177–EL183, Feb. 2020, doi: 10.1121/10.0000749.

[4] Z.-J. Zhu, S.-Q. Ma, X.-Q. Zhu, Q. Lan, S.-C. Piao, and Y.-S. Cheng, "Parallel optimization of underwater acoustic models: A survey," *Chin. Phys. B*, vol. 31, no. 10, Oct. 2022, Art. no. 104301, doi: 10.1088/1674-1056/ac7ccc.

[5] P. Petrov, B. Katsnelson, and Z. Li, "Modeling techniques for underwater acoustic scattering and propagation (including 3D effects)," *J. Mar. Sci. Eng.*, vol. 10, no. 9, p. 1192, Aug. 2022. [Online]. Available: https://www.mdpi.com/2077-1312/10/9/1192

[6] Y. Chen, Q. Ma, J. Yu, and T. Chen, "Underwater acoustic object discrimination for few-shot learning," in *Proc. 4th Int. Conf. Mech., Control Comput. Eng. (ICMCCE)*, Oct. 2019, pp. 430–4304. [Online]. Available: https://ieeexplore.ieee.org/document/8969465

[7] D. Li, F. Liu, T. Shen, L. Chen, and D. Zhao, "Data augmentation method for underwater acoustic target recognition based on underwater acoustic channel modeling and transfer learning," *Appl. Acoust.*, vol. 208, Jun. 2023, Art. no. 109344. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X23001421

[8] Q. Yao, Y. Wang, and Y. Yang, "Underwater acoustic target recognition based on data augmentation and residual CNN," *Electronics*, vol. 12, no. 5, p. 1206, Mar. 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/5/1206

[9] G. Jin, F. Liu, H. Wu, and Q. Song, "Deep learning-based framework for expansion, recognition and classification of underwater acoustic signal," *J. Experim. Theor. Artif. Intell.*, vol. 32, no. 2, pp. 205–218, Mar. 2020, doi: 10.1080/0952813x.2019.1647560.

[10] Y. Chen, H. Lu, J. Yu, and H. Wang, "Improved method for generative adversarial nets," in *Proc. Int. Conf. Intell. Comput. Hum.-Comput. Interact. (ICHCI)*, Dec. 2020, pp. 404–408. [Online]. Available: https://ieeexplore.ieee.org/document/9424821

[11] H. I. Hummel, R. van der Mei, and S. Bhulai, "A survey on machine learning in ship radiated noise," *Ocean Eng.*, vol. 298, Apr. 2024, Art. no. 117252. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801824005894

[12] J. D. 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, p. 2648, Jun. 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/11/2648

[13] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-López, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Appl. Acoust.*, vol. 113, pp. 64–69, Dec. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X16301566

[14] 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, Nov. 2021, Art. no. 115270. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421007016

[15] X. Du and F. Hong, "Introducing the brand new QiandaoEar22 dataset for specific ship identification using ship-radiated noise," 2024, *arXiv:2406.04353*.

[16] A. Da Silva Borges and S. Figueira-Cardoso, "A common archetype: Imaginary and linguistic-discursive analysis of the heroic feminine in Brazilian and Polish folk narratives," *Ameryka Lacinska Kwartalnik Analityczno-Informacyjny*, vol. 1, no. 117, pp. 101–122, Nov. 2022. [Online]. Available: http://amerykalacinska.com/resources/html/article/details?id=233381

[17] N. Nandakumaran, R. Tharmarasa, T. Lang, and T. Kirubarajan, "Gaussian mixture probability hypothesis density smoothing with multistatic sonar," *Proc. SPIE*, vol. 6968, pp. 115–122, Apr. 2008. [Online]. Available: https://api.semanticscholar.org/CorpusID:129424381

[18] W. Shin, D.-S. Kim, and H. Ko, "Target tracking from weak acoustic signals in an underwater environment using a deep segmentation network," *J. Mar. Sci. Eng.*, vol. 11, no. 8, p. 1584, Aug. 2023. [Online]. Available: https://www.mdpi.com/2077-1312/11/8/1584

[19] S. Cominelli, N. Bellin, C. D. Brown, V. Rossi, and J. Lawson, "Acoustic features as a tool to visualize and explore marine soundscapes: Applications illustrated using marine mammal passive acoustic monitoring datasets," *Ecol. Evol.*, vol. 14, no. 2, p. 10951, Feb. 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.10951

[20] W. S. Filho, J. M. de Seixas, and N. N. de Moura, "Preprocessing passive sonar signals for neural classification," *IET Radar, Sonar Navigat.*, vol. 5, no. 6, pp. 605–612, Jul. 2011. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-rsn.2010.0157

[21] H. Kosarirad, M. G. Nejati, A. Saffari, M. Khishe, and M. Mohammadi, "Feature selection and training multilayer perceptron neural networks using grasshopper optimization algorithm for design optimal classifier of big data sonar," *J. Sensors*, vol. 2022, pp. 1–14, Nov. 2022. [Online]. Available: https://www.hindawi.com/journals/js/2022/9620555/

[22] X. Yin, X. Sun, P. Liu, L. Wang, and R. Tang, "Underwater acoustic target classification based on LOFAR spectrum and convolutional neural network," in *Proc. 2nd Int. Conf. Artif. Intell. Adv. Manuf.*, New York, NY, USA, Oct. 2020, pp. 59–63, doi: 10.1145/3421766.3421890.

[23] Q. Meng and S. Yang, "A wave structure based method for recognition of marine acoustic target signals," *J. Acoust. Soc. Amer.*, vol. 137, no. 4, p. 2242, Apr. 2015. [Online]. Available: https://pubs.aip.org/jasa/article/137/4Supplement/2242/709514/A-wave-structure-based-method-for-recognition-of

[24] X. Jiang, Q. Wang, and X. Zeng, "Cavitation noise classification based on spectral statistic features and PCA algorithm," in *Proc. 3rd Int. Conf. Comput. Sci. Netw. Technol.*, Oct. 2013, pp. 438–441. [Online]. Available: https://ieeexplore.ieee.org/document/6967148

[25] S. Li, X. Jin, S. Yao, and S. Yang, "Underwater small target recognition based on convolutional neural network," in *Proc. Global Oceans*, Oct. 2020, pp. 1–7. [Online]. Available: https://ieeexplore.ieee.org/document/9389160

[26] O. Axelsson and C. Rhén, "Neural-network-based classification of commercial ships from multi-influence passive signatures," *IEEE J. Ocean. Eng.*, vol. 46, no. 2, pp. 634–641, Apr. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9123373

[27] B. Beckler, A. Pfau, M. Orescanin, S. Atchley, N. Villemez, J. E. Joseph, C. W. Miller, and T. Margolina, "Multilabel classification of heterogeneous underwater soundscapes with Bayesian deep learning," *IEEE J. Ocean. Eng.*, vol. 47, no. 4, pp. 1143–1154, Oct. 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9832855

[28] X. Wang, C. Zhang, and R. Wang, "Underwater acoustic target recognition technology based on MFA-conformer," in *Proc. 2nd Int. Conf. Electron. Inf. Eng. Comput. Technol. (EIECT)*, Oct. 2022, pp. 214–217. [Online]. Available: https://ieeexplore.ieee.org/document/10066554

[29] Y. Yao, X. Zeng, H. Wang, and J. Liu, "Research on underwater acoustic target recognition method based on DenseNet," in *Proc. 3rd Int. Conf. Big Data, Artif. Intell. Internet Things Eng. (ICBAIE)*, Jul. 2022, pp. 114–118. [Online]. Available: https://ieeexplore.ieee.org/document/9985924

[30] J. Yu, P. Wang, Z. Cai, and J. Shang, "Underwater acoustic target classification and grading technology based on track fusion," in *Proc. 2nd Int. Symp. Sensor Technol. Control (ISSTC)*, Aug. 2023, pp. 144–150. [Online]. Available: https://ieeexplore.ieee.org/document/10281199

[31] F. Liu, T. Shen, Z. Luo, D. Zhao, and S. Guo, "Underwater target recognition using convolutional recurrent neural networks with 3-D mel-spectrogram and data augmentation," *Appl. Acoust.*, vol. 178, Jul. 2021, Art. no. 107989. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003682X21000827

[32] H. Ashraf, B. Shah, A. M. Soomro, Q.-U.-A. Safdar, Z. Halim, and S. K. Shah, "Ambient-noise free generation of clean underwater ship engine audios from hydrophones using generative adversarial networks," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107970. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790622002452

[33] S. Yang, A. Jin, X. Zeng, H. Wang, X. Hong, and M. Lei, "Underwater acoustic target recognition based on sub-band concatenated mel spectrogram and multidomain attention mechanism," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 107983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197624001416

[34] *Technical Characteristics for an Automatic Identification System Using Time Division Multiple Access in the VHF Maritime Mobile Frequency Band*, document ITU-R M.1371-5, Int. Telecommun. Union (ITU), Feb. 2014. [Online]. Available: https://www.itu.int/dmspubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf

[35] Y. Li, Y. Zhou, and S. Jiao, "Variable-step multiscale Katz fractal dimension: A new nonlinear dynamic metric for ship-radiated noise analysis," *Fractal Fractional*, vol. 8, no. 1, p. 9, Dec. 2023. [Online]. Available: https://www.mdpi.com/2504-3110/8/1/9

[36] J. Ren, Y. Xie, X. Zhang, and J. Xu, "UALF: A learnable front-end for intelligent underwater acoustic classification system," *Ocean Eng.*, vol. 264, Nov. 2022, Art. no. 112394. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801822016833

[37] S. Feng and X. Zhu, "A transformer-based deep learning network for underwater acoustic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9866066

[38] R. P. Hodges, *Underwater Acoustics: Analysis, Design, and Performance of Sonar*, 1st ed., Hoboken, NJ, USA: Wiley, 2010.

[39] H. Wu, Q. Song, and G. Jin, "Underwater acoustic signal analysis: Preprocessing and classification by deep learning," *Neural Netw. World*, vol. 30, no. 2, pp. 85–96, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:129424381

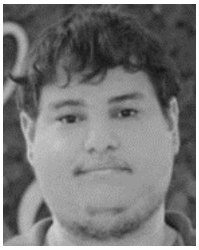
[40] A. Zhou et al., "A novel cross-attention fusion-based joint training framework for robust underwater acoustic signal recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4209516, doi: 10.1109/TGRS.2023.3333971.

[41] B. Snapir, T. Waine, and L. Biermann, "Maritime vessel classification to monitor fisheries with SAR: Demonstration in the north sea," *Remote Sens.*, vol. 11, no. 3, p. 353, Feb. 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/3/353

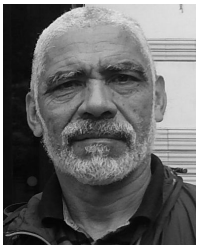
[42] E. Alpaydm, "Combined 5 × 2 cv F test for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 11, no. 8, pp. 1885–1892, Nov. 1999, doi: 10.1162/089976699300016007.

[43] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manage.*, vol. 45, no. 4, pp. 427–437, Jul. 2009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306457309000259

FÁBIO OLIVEIRA BAPTISTA DA SILVA received the degree in electronic and computer engineering from the Federal University of Rio de Janeiro (UFRJ), in 2014, and the M.S. degree in electrical engineering from the Electrical Engineering Program (PEE), Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE), UFRJ, in 2020, where he is currently pursuing the Ph.D. degree in electrical engineering. Since 2016, he has been a Military Officer and a Researcher with the Submarine Acoustic Systems Group, Navy Research Institute. He has worked on various projects related to calibration, instrumentation, signal processing, and developing systems for active and passive sonars.



JÚLIO DE CASTRO VARGAS FERNANDES received the Ph.D. degree in electrical engineering from the Electrical Engineering Program (PEE), Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE), Federal University of Rio de Janeiro (UFRJ). He is currently a Researcher at the Laboratório Nacional de Computação Científica (LNCC) (in english, National Laboratory for Scientific Computing), focusing on machine learning performing research. He has been researching applied machine learning in several fields, such as oil and gas, defense, and digital signal processing. His research interests include neural networks, deep learning, ensemble learning, signal processing, and computational intelligence.



WILLIAM SOARES FILHO received the B.S. degree in electronic and telecommunications engineering from the Pontifical Catholic University of Minas Gerais, in 1978, and the M.S. degree in biomedical engineering and the Ph.D. degree in electrical engineering from the Federal University of Rio de Janeiro (UFRJ), in 1982 and 2001, respectively. He has extensive experience in electrical engineering, with a focus on acoustic signal processing. His research interests include digital signal processing, passive sonar, neural networks, underwater acoustic signal processing, and contact tracking in passive and active sonar systems.

JOÃO BAPTISTA DE OLIVEIRA E SOUZA FILHO received the degree in electrical engineering with an emphasis in electronics from the Escola Politécnica da Universidade Federal do Rio de Janeiro (POLI), Federal University of Rio de Janeiro (UFRJ), in 2001, and the M.Sc. and D.Sc. degrees in electrical engineering from the Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE), Federal University of Rio de Janeiro (UFRJ), in 2002 and 2007, respectively. He is currently a Full Professor with the Department of Electronic and Computer Engineering (DEL), POLI, UFRJ, and a Full Faculty Member of the Graduate Program in Electrical Engineering, Electrical Engineering Program (PEE), COPPE, UFRJ. He co-authored the monograph *Online Component Analysis, Architectures, and Applications* (Delft: NOW Publishers, 2022, with P. S. R. Diniz, L. D. Van, and T. P. Jung). His teaching and research interests include machine learning, deep learning, natural language processing, online learning, digital signal processing, embedded systems, and electronic instrumentation, with significant contributions to multidisciplinary projects in the areas of defense, medicine, oil and gas, climate, and multimedia. He also serves as an Associate Editor for IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS.

ÂNGELA SPENGLER received the degree in oceanology from the Federal University of Rio Grande (FURG), in 2005, and the M.S. degree in oceanography from the Federal University of Pernambuco (UFPE), in 2009. She has been working in environmental monitoring at Petrobras: Brazilian O&G Company, since 2011, and currently coordinates the Santos Basin Soundscape Monitoring Project (PMPAS-BS) among other projects.

NATANAEL NUNES DE MOURA JÚNIOR received the degree in electronic and computer engineering, the M.S. degree in electrical engineering, and the Ph.D. degree in electrical engineering from the Federal University of Rio de Janeiro (UFRJ), in 2011, 2013, and 2018, respectively. He is currently a Professor at UFRJ and a Full Member of the Signal Processing Laboratory. He reviews several journals, including *IET Radar, Sonar and Navigation*, IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, and IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS. His expertise lies in electrical engineering, with a focus on signal processing. His main research interests include computational intelligence, machine learning, and passive sonar systems.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - ROR identifier: 00x0ma614