# From Backscattering Models to Neural Networks: Simulation-Driven Acoustic Identification of Mesopelagic Organisms

Sander Andre Berg Marx[1,*], Babak Khodabandeloo[2,†], Ketil Malde[2,‡]
[1]Department of Informatics, University of Bergen, Bergen, Norway
[2]Acoustics and Observation Methodologies, Institute of Marine Research (IMR), Bergen, Norway
[*]e-mail: sanderamarx@hotmail.com
[†]e-mail: babak.khodabandeloo@hi.no
[‡]e-mail: ketil.malde@hi.no

***Abstract*—Target Strength (TS) is the logarithmic measure of the backscattered acoustic energy reflected toward a sound source when an acoustic wave encounters an organism. It depends on the organism's size, shape, material properties, orientation, and the frequency of the wave, and thus carries information useful for identifying and characterizing marine organisms. Advanced broadband echosounders now allow detailed TS measurements of marine organisms over nearly continuous frequency ranges, which provide valuable information for biomass estimation and ecosystem monitoring. However, interpreting these TS measurements has traditionally relied on manual classification, which makes it difficult to extract biological characteristics for target classification or ecological analysis, especially given the complexity of broadband data. Physics-based backscattering models are versatile tools for modeling the TS frequency response given the shape and material properties of the scatterers. In this study, we employ an exact prolate spheroid model, representative of many marine organisms, to simulate broadband TS spectra for training machine learning models. These models aim to classify and characterize targets based on their TS frequency signatures. A hybrid one-dimensional Convolutional Neural Network (1D-CNN) is proposed for the simultaneous classification (gas- vs. liquid-filled) and regression of geometric properties, and compared against K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). Furthermore, the model's generalization capability is evaluated against *in situ* broadband measurements of mesopelagic organisms. Results show that while all models achieved perfect classification accuracy, the hybrid 1D-CNN clearly outperformed others in parameter estimation. When applied to real-world data, the model achieved high precision in size estimation ($R_{\mathrm{eq}}$), though challenges remain in predicting elongation. By feeding the 1D-CNN's predictions back into the prolate spheroid model, we observe that the model accurately predicts the frequency of the first resonance peak but struggles with the amplitude and location of the second resonance peak. This discrepancy is attributed to physical damping effects present in real biological tissues but absent in the simulated training data. This demonstrates that simulation-driven machine learning is a promising strategy for overcoming data scarcity, though incorporating realistic biological damping and real-world noise effects remains a critical step toward fully automated acoustic identification.**

***Keywords-Acoustic target classification; machine learning; Convolutional neural network; prolate spheroid backscattering model; simulation-driven learning; in situ validation; real-world acoustic data.***

## I. Introduction

This article is an extended version of the COCE 2025 conference paper "Machine Learning Using Physics-Based Backscattering Modeling for Acoustic Identification of Mesopelagic Organisms" [1]. Compared with the original conference paper, this journal submission expands the study by adding validation of the model performance on *in situ* real-world acoustic data. Active acoustics is a versatile tool for monitoring marine life, offering unrivaled spatial and temporal resolution compared to other methods, such as net-based biological sampling, optical systems, or video recording. However, interpreting the collected echoes to classify organism size and species remains challenging. Converting acoustic data to biomass is especially difficult when the insonified volume contains mixed species and/or a diverse size distribution within the same species [2][3][4]. Broadband echosounders provide detailed spectra, but these are difficult to interpret manually, making large-scale analysis time-consuming and thereby motivating automated approaches.

When an acoustic wave encounters organisms along its propagation path, the acoustic energy scatters in different patterns depending on the organism's size, shape, material properties, orientation, and the frequency of the wave [5][6]. A portion of this energy is reflected back toward the source, referred to as backscattering. Therefore, the backscattered signal—or its logarithmic measure, known as Target Strength (TS)—carries information that can be used to identify the object from which it originated. Since the TS of an organism varies with frequency, measuring TS over a broad, continuous frequency band provides more detailed information about the scatterers, thereby enhancing the ability to characterize organisms [7]. Broadband fisheries echosounders transmit frequency-modulated pulses, offering two main advantages [8][9]: (1) measurement across a nearly continuous frequency range, and (2) enhanced range resolution through pulse compression, which improves single-target detection. Therefore, the use of broadband echosounders improves the ability to identify organisms and to resolve their spatial distribution within aggregations, both of which are crucial for interpreting and converting acoustic echoes into meaningful biological information.

Backscattering models are an essential tool for interpreting measured backscattered data from marine organisms [10]. Techniques to model acoustic backscattering range from analytical to numerical models. Numerical models can accommodate arbitrary geometries and inhomogeneous material properties, allowing detailed and realistic simulations [11]. However, their high computational cost can limit their practical use, especially when generating massive datasets required to train modern deep learning architectures [12]. To make simulation-driven machine learning feasible, optimizing the execution time of these physics-based models is a critical prerequisite. Although most marine organisms have complex geometries and consist of inhomogeneous materials, backscattering models from canonical geometries with homogeneous material properties are often sufficient to acoustically represent them [13]. For example, spherical models have long been used in fisheries acoustics due to their simplicity [14][15]. However, spherical models cannot capture elongation, which is a critical parameter for accurately representing many marine organisms or their main acoustic reflecting organ, e.g., gas-bladder.

Machine Learning (ML) techniques can be used to classify targets based on trained models that learn to find patterns within the broadband TS spectra. These approaches are generally categorized into unsupervised methods, which group targets based on spectral similarity [16], and supervised methods, which map acoustic signatures to specific biological classes or physical parameters. Supervised learning is typically the most successful approach for precise identification, but it requires large amounts of accurately labeled data to develop robust models [12].

However, acquiring such labels, referred to as "ground truth," is exceptionally challenging for individual fish in the wild. The acoustic scattering environment is highly dynamic; internal structures such as bones, flesh, and the swimbladder create complex impedance contrasts, while constructive and destructive interference across the pulse bandwidth varies significantly as the fish moves. While some studies, such as Kubilius et al. [17], have established ground truth in controlled settings using artificial fish-like targets of known dimensions, obtaining reliable labels for wild organisms remains difficult. Conventional methods rely on invasive biological sampling, such as trawling, which introduces inherent catch biases and provides only a population-level snapshot rather than individual-target validation [17][18].

Furthermore, while real-world broadband acoustic datasets exist, such as the *in situ* measurements used in this study, they are generally limited in scale. These datasets often lack direct physical measurements of the targets and must instead rely on indirect analytical estimations—such as resonance frequency analysis to provide labels. Such estimations, while physically grounded, do not provide the high-confidence ground truth required to train modern, data-intensive deep learning architectures [12].

To overcome these limitations of data scarcity and label uncertainty, we use a physics-based simulation model [19] to generate large labeled synthetic datasets. This paper presents a two-fold contribution: First, we implement and evaluate a computationally optimized version of a prolate spheroid backscattering model to enable efficient data generation. Second, we propose a novel hybrid one-dimensional Convolutional Neural Network (1D-CNN) for the simultaneous classification (gas-filled vs. liquid-filled) and regression of geometric properties such as size, elongation, and orientation.

The performance of the 1D-CNN is then compared against traditional ML models. The method is also tested on a small *in situ* broadband dataset of mesopelagic organisms, specifically small gas-bearing fish, from Khodabandeloo et al. [20]. This allows us to evaluate the model's ability to generalize from idealized simulations to noisier, real-world acoustic measurements.

The remainder of the paper is organized as follows. Section II describes the methods for computational optimization, data generation, and machine learning model training. Section III reports the results on both synthetic and *in situ* data, and Section IV provides a discussion of the findings, including the challenges of real-world generalization. Section V concludes the paper and outlines directions for future work.

## II. Methods

### A. Physics-Based Acoustic Backscattering Modeling for Synthetic Data Generation

To generate synthetic broadband backscattering data from objects representing mesopelagic marine organisms, with and without gas bladders, we employed an optimized version of the fluid-filled prolate spheroid backscattering model [19]. This physics-based model provides accurate backscattering for liquid- and gas-filled targets over a wide frequency range and for all incident angles, i.e., the relative orientation of the prolate spheroid axis to the direction of wave propagation.

The prolate spheroids are represented by their volume (quantified by the equivalent spherical radius $\boldsymbol{R}_{\mathbf{eq}}$) and elongation (or aspect ratio $\alpha$) instead of directly using semi-major and semi-minor axes, a and b, respectively. They are related by the following equations:

$$b = \frac{R_{\mathrm{eq}}}{\alpha^{1/3}}, \tag{1}$$

$$a = \alpha \times b. \tag{2}$$

This parameterization facilitates direct control over the size and shape of the modeled organisms, making it easier to explore a wide range of biologically relevant geometries. The range of parameters used to model different organisms are given in Table I.

TABLE I. Parameter ranges used for generating synthetic acoustic data

| Parameter | Targets | |
|---|---|---|
| | **Gas-filled** | **Liquid-filled** |
| Equivalent radius ($R_{\mathrm{eq}}$) | 0.1–5 mm | 5–20 mm |
| Aspect ratio ($\alpha$) | 1.05–8.0 | 1.05–8.0 |
| Spheroid density ($\rho_s$) | 2–80 kg/m$^3$ | 1.01–1.07 $\times\rho_w$ kg/m$^3$ |
| Water density ($\rho_w$) | 1027 kg/m$^3$ | 1027 kg/m$^3$ |
| Spheroid sound speed ($c_s$) | 343 m/s | 1.01–1.07 $\times c_w$ m/s |
| Water sound speed ($c_w$) | 1500 m/s | 1500 m/s |
| Incident angle ($\theta$) | 0.01–90° | 0.01–90° |
| Frequency range | 10–260 Hz (0.5 kHz steps) | |

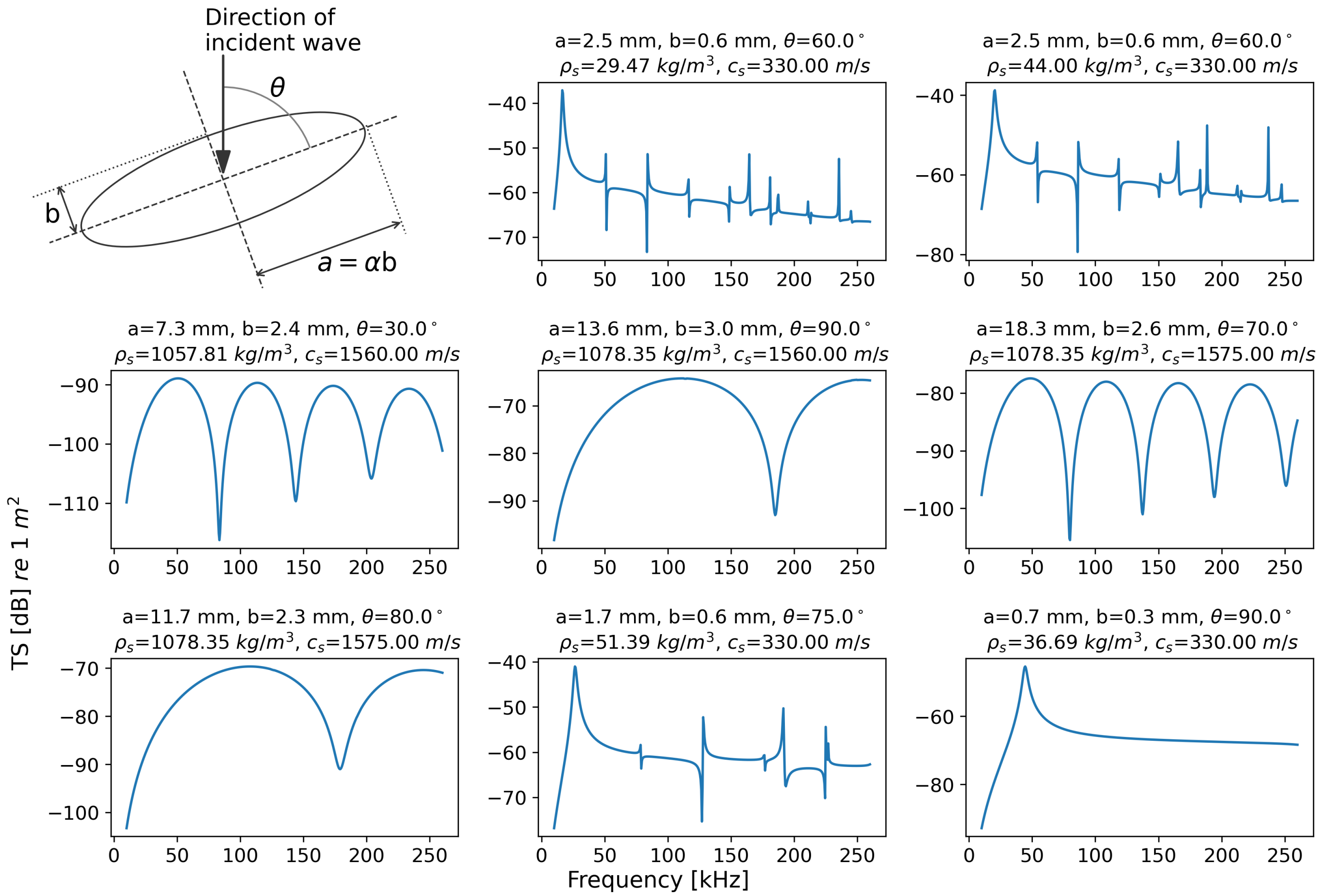


Figure 1. Examples of TS(f) for prolate spheroids with parameters within the ranges in Table I. An illustration of a prolate spheroid with semi-major and semi-minor axes, a and b respectively, along with the incident angle $\theta$, is shown in the upper left corner.

### *B. Computational Optimization*

To enable the generation of large-scale datasets, the original simulation code, a hybrid of Python and Fortran was refactored for efficiency. We performed code profiling to identify computational bottlenecks, which were primarily located in nested Python loops managing frequency-point calculations. To resolve this, we implemented NumPy vectorization, replacing iterative Python loops with array-based operations optimized in C. This approach allowed for the simultaneous calculation of entire frequency bands rather than sequential point-by-point processing. Furthermore, the optimized code was designed to run in parallel across 200 CPU cores on a high-performance computing cluster to maximize data throughput as the frequency-point calculations are independent.

### *C. Data Preprocessing and Machine Learning Setup*

The synthetic acoustic data generated by the prolate spheroid backscattering model underwent several preprocessing steps to prepare it for machine learning. The dataset consists of 30,000 simulated TS spectra, 15,000 liquid-filled and 15,000 gas-filled targets.

Each target is represented by a feature vector with corresponding labels. The input features consist of the broadband TS spectra. Each spectrum is represented as a vector of TS values, sampled at 0.5 kHz steps across the frequency range of 10 to 260 kHz. The models were trained to predict a classification label (liquid-filled or gas-filled) and three regression labels corresponding to the geometric properties of the prolate spheroid: the incident angle ($\theta$), the semi-major axis (a), and the semi-minor axis (b).

Normalization procedures are applied separately to the feature and label columns to ensure consistent scaling and improve model stability [21]. Each frequency TS value is standardized independently using training set statistics, transforming the data to have zero mean and unit variance. The continuous regression labels ($\theta$, a and b) are also normalized to ensure that each parameter contributes equally to model training regardless of original scale.

The complete dataset of 30,000 target instances is then partitioned into training, validation, and test subsets using an 80/10/10 split. The training set (80%) is used for training the ML models and deriving normalization statistics. The validation set (10%) is used for hyperparameter tuning and model selection during training, as well as for monitoring overfitting. Finally, the test set (10%) is reserved for the final, unbiased evaluation of the trained models' performance.

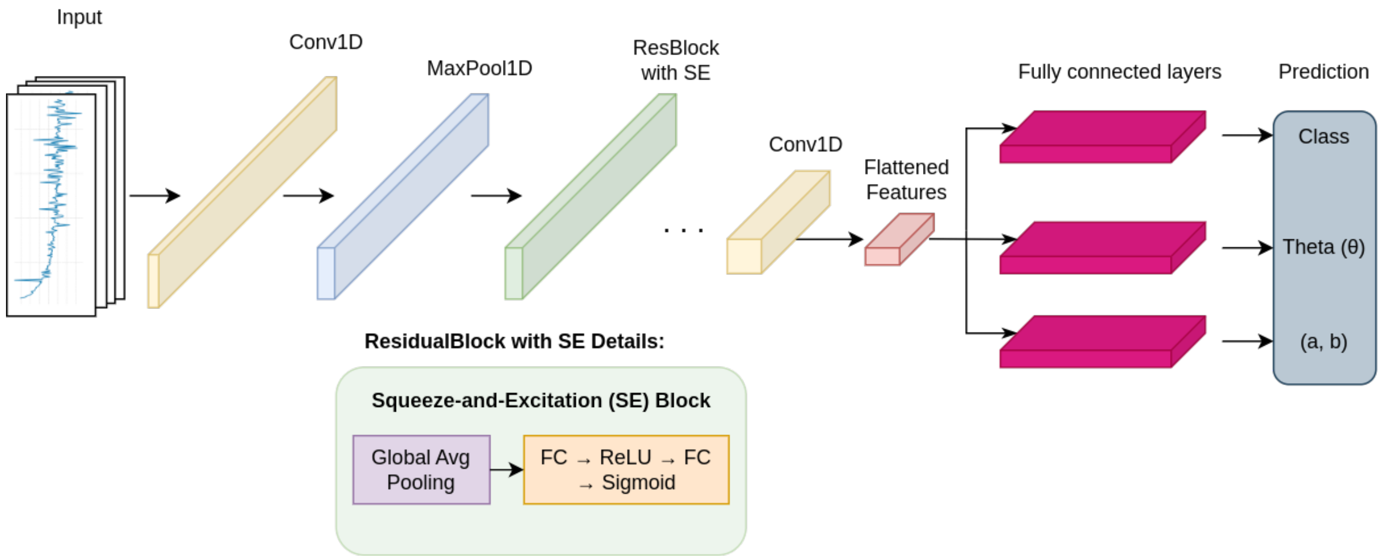


Figure 2. Hybrid one-dimensional convolutional neural network architecture for acoustic target classification.

### *D. Machine Learning Models*

To classify the type and estimate geometric parameters from the TS spectra, a hybrid 1D-CNN was developed. The architecture was chosen for its effectiveness in automatically extracting hierarchical features from sequential data, such as acoustic frequency spectra [22].

The network architecture, illustrated in Figure 2, was designed to process sequential TS data. The architecture incorporates advanced deep learning components, specifically residual blocks [23] and Squeeze-and-Excitation (SE) attention mechanisms [24], to improve feature extraction and training stability. The TS spectra are first processed through an initial 1D convolution layer, followed by a series of residual blocks. The residual blocks contain two 1D convolutional layers, batch normalization [25], and ReLU activations. Squeeze-and-Excitation (SE) attention mechanisms, which perform channel-wise feature weighting by compressing information through global average pooling and then learning channel relationships via two fully connected layers, are also integrated into these residual blocks. The network ends in fully connected layers that lead to a multi-head output structure with dedicated heads for classification and regression: one head for target-type classification, one for regression of the incident angle ($\theta$), and one for regression of the size parameters (a and b).

For comparative analysis of the performance of the 1D-CNN, three additional traditional machine learning models previously used in acoustic target classification research were tested.

- K-Nearest Neighbors (KNN) [26]: An instance-based learner classifying targets based on the majority class of their k nearest neighbors in the feature space, as applied by Cotter et al. [27].
- Support Vector Machine (SVM) [28]: Seeking an optimal hyperplane to separate classes or predict continuous values, utilized by Yang et al. [29] for underwater target recognition.
- Random Forest (RF) [30]: An ensemble method constructing multiple decision trees on random subsets of data and features, aggregating their predictions, as employed by Gugele et al. [31].

These traditional models were not hyperparameter tuned and served as a baseline to evaluate the 1D-CNN.

### *E. In Situ Generalization Setup*

To evaluate real-world performance, a second experiment was designed using an *in situ* broadband dataset of nine single mesopelagic organisms [20], interpreted as swimbladdered mesopelagic fish (gas-bearing targets) based on their characteristic resonance structure. Because the targets were identified acoustically, species-level labels were not available; we therefore treat them as gas-bearing organisms dominated acoustically by a swimbladder-like gas inclusion.

The simulation parameters were adjusted to match the characteristics of these targets. The training dataset was constrained to gas-filled targets with a semi-major axis $a \leq$ 1 mm. To account for the depth-dependent gas compression, the spheroid density ($\rho_s$) was calculated using the Van der Waals equation [32][33] based on the target depths (140 to 575 meters). Gaussian noise was added to the synthetic training data to replicate environmental and instrumental noise present in the *in situ* data. Finally, the frequency range was adjusted to 38–380 kHz, including gaps to replicate the echosounder channels used during the *in situ* collection.

### *F. Training and Evaluation*

The 1D-CNN was trained using a custom multi-component loss function designed to address both the classification and regression tasks. The total loss ($\mathcal{L}$) is a weighted sum of four components, as shown in (3):

$$\mathcal{L} = \omega_c \, \mathcal{L}_\text{c} + \omega_r \left( \mathcal{L}_\theta + \mathcal{L}_{ab} \right) + \omega_{cons} \, \mathcal{L}_\text{cons} \qquad (3)$$

where $\mathcal{L}_c$ is cross-entropy loss for the binary classification task. For regression, Huber loss [34] was used for the angle ($\mathcal{L}_\theta$) and Mean Squared Error (MSE) for the size parameters ($\mathcal{L}_{ab}$). To enforce the geometric relationship between the semi-major and minor axes, a consistency loss ($\mathcal{L}_{\text{cons}}$), also based on Huber loss, was applied by comparing the predicted aspect ratio ($\alpha_{pred} = a_{pred}/b_{pred}$) to its ground truth value. The weights were determined through hyperparameter tuning and set to $\omega_c$ = 0.1, $\omega_r$ = 2.0, and $\omega_{cons}$ = 0.5 to prioritize the more challenging regression task.

Model performance was evaluated on the unseen test set. Classification performance was measured by accuracy, while regression performance was assessed using the coefficient of determination ($R^2$) and Root Mean Squared Error (RMSE) for $\theta$, a, b, and the derived $\alpha$. A composite score, averaging the classification accuracy and the $R^2$ scores of the three primary regression targets, was used for overall model comparison.

## III. Results

### A. Optimization Results

Before training the ML models, the performance of the NumPy-vectorized scattering model was benchmarked against the original implementation. A “frequency point” is defined as a single calculation of the backscattering cross-section at a specific frequency; a full TS spectrum in this study consists of over 500 such points.

To generate the 39,400 spectra required for this study, the optimized code was executed in parallel across 200 CPU cores on the “Birget” cluster, for a total wall-clock time of 16 days. This equates to approximately 77,000 total CPU hours. Based on the benchmarked 47.6-fold speedup, which reduced the average computation time per frequency point from 334.94 seconds to 7.04 seconds, we can estimate the impact of this optimization on the project timeline.

Using the same 200-core parallel infrastructure, the non-optimized implementation would have required an estimated 760 days of continuous cluster usage to generate the same dataset. This estimate is approximate, as the computational cost varies with target parameters (e.g., size, orientation, and whether the inclusion is gas- or liquid-filled).

The original implementation is used here as a benchmarking reference to quantify relative performance improvements, rather than as an absolute measure of computational efficiency under all conditions. Even allowing for variability, the results clearly indicate that the NumPy-vectorized optimization was a necessary prerequisite for generating the 39,400 spectra within a realistic research timeframe. For more details of the benchmark, see Figure 17 in Khodabandeloo et al. [19].

### B. Performance Comparison Across Models (Synthetic data)

Using the simulated backscattering frequency responses, we first classified the targets (i.e., liquid- or gas-filled), then estimated the geometrical properties (semi-major and semi-minor axes) and the incident angle of the prolate spheroids based on their wideband target strength frequency responses, using different ML models. Since the data are simulated, the true model parameters are known, allowing us to quantitatively evaluate the estimates produced by the various ML models. The performance of the models in both classification and morphological parameter estimation is summarized in Table II.

TABLE II. Comparative regression performance ($R^2$ scores), classification accuracy, and composite score on synthetic test data

| Parameter | SVM | KNN | RF | 1D-CNN |
|---|---|---|---|---|
| $R^2$ for $\theta$ | 0.742 | 0.751 | 0.853 | **0.974** |
| $R^2$ for $a$ | 0.795 | 0.898 | 0.902 | **0.997** |
| $R^2$ for $b$ | 0.911 | 0.931 | 0.942 | **0.996** |
| Clas. Acc. | 1.000 | 1.000 | 1.000 | **1.000** |
| Composite score | 0.719 | 0.804 | 0.867 | **0.992** |

### C. Detailed Performance of Hybrid 1D-CNN

All tested models achieved a perfect classification accuracy of 1.00, correctly identifying every target as either gas-filled or liquid-filled. The primary differences between the models emerged in the regression tasks. Given its superior regression performance, the results of the hybrid 1D-CNN were analyzed in further detail. Figure 3 presents scatter plots of the model's predicted values against the true values for each regression target.

The model demonstrated high precision in predicting the semi-major axis a and semi-minor axis b; data points were tightly clustered along the identity line for both gas-filled and liquid-filled targets, indicating the model's robustness across target types and size ranges. Predictions for the incident angle $\theta$ were also strong, though with more visible scatter compared to the size parameters a and b, with notable gas-filled outliers. The derived aspect ratio ($\alpha = a/b$) also showed good performance, but with some increased scatter for targets with higher aspect ratios ($\alpha$ > 4).

### D. Analysis of Prediction Outliers

Outlier analysis revealed consistent patterns for both incident angle ($\theta$) and aspect ratio ($\alpha$).

For incident angle ($\theta$) estimation, the largest errors occurred with small, gas-filled targets with simple TS spectra containing few distinct resonance peaks. The model therefore defaulted toward predicting the training mean (45°), creating the horizontal band of outliers seen in Figure 3. In contrast, liquid-filled outliers showed smaller angle errors and were typically targets with low aspect ratio ($\alpha \approx 1.05$), where the orientation is less defined physically as the target approaches a perfect sphere.

A similar pattern emerged for aspect ratio ($\alpha$) prediction, where most outliers were again small, gas-filled targets with spectrally simple signatures. This result aligns with the physical principle that accurate estimation of elongation requires at least two distinct resonance peaks [35], a feature these outlier spectra lacked.

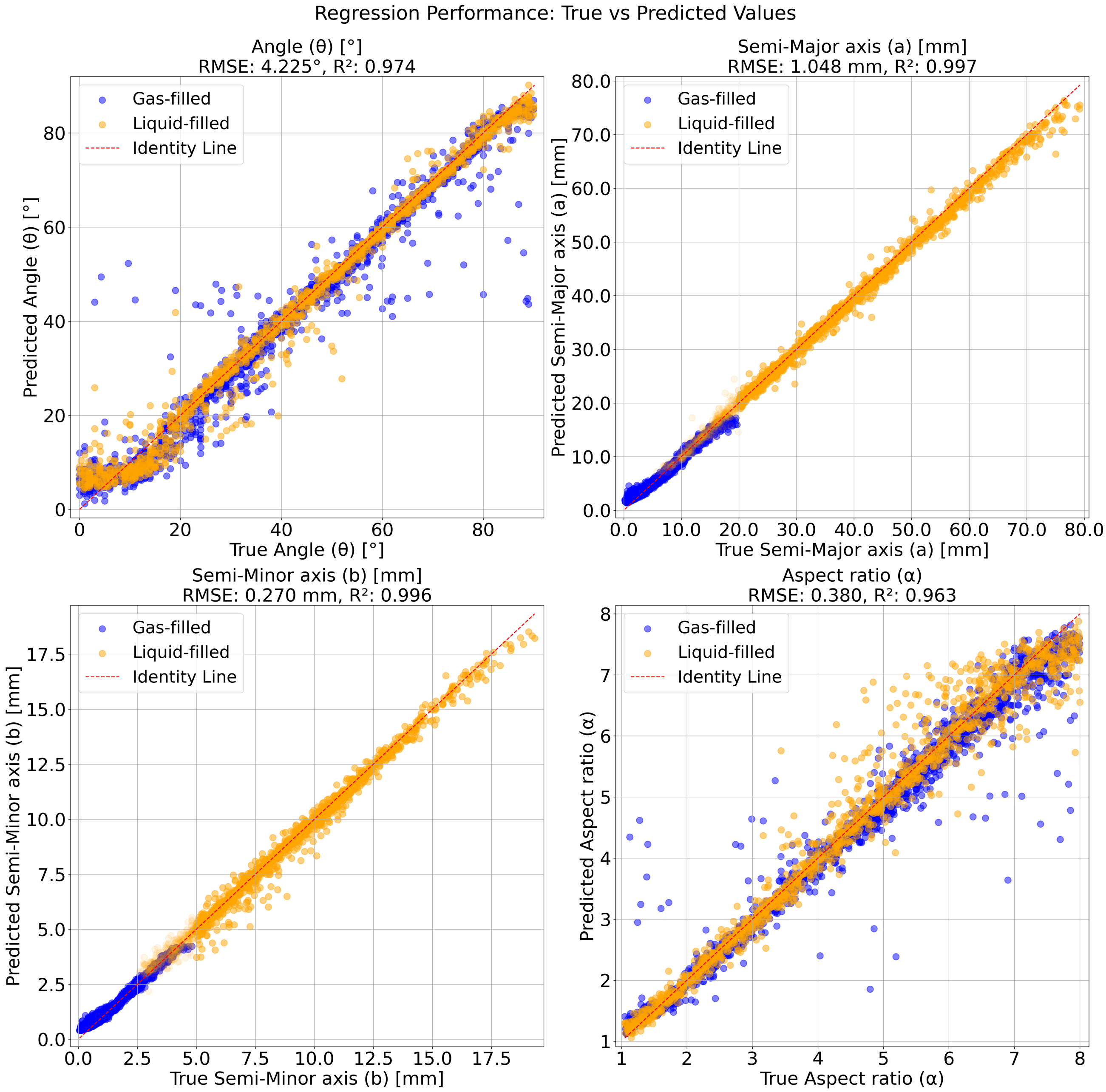


Figure 3. Regression performance of the hybrid 1D-CNN model on test data showing true values against predicted values for (top-left) angle ($\theta$), (top-right) semi-major axis (a), (bottom-left) semi-minor axis (b), and (bottom-right) derived aspect ratio ($\alpha = a/b$). Blue points represent gas-filled targets, and orange points represent liquid targets.

### *E. Performance of In Situ Measurements*

To test real-world generalization, the 1D-CNN trained on simulated data was applied to twelve *in situ* acoustic targets. For the first nine targets, the first and second resonance peaks, as well as corresponding estimates of size and elongation, were obtained from Khodabandeloo et al. [20]. The model's predictions of equivalent radius ($R_{eq}$) and aspect ratio ($\alpha$) were then compared against values derived from resonance analysis on targets 1-9 [20], as detailed in Table III.

The model demonstrated high accuracy in predicting the equivalent radius ($R_{eq}$), achieving a Mean Absolute Error (MAE) of only 0.042 mm. Target 3, for instance, showed a radius error of 0.000 mm, while the largest error (Target 6) was limited to 0.106 mm. Aspect ratio ($\alpha$) estimation was more variable, with an MAE of 0.94; Target 7 showed the highest accuracy ($|\varepsilon_\alpha| = 0.14$), while Target 9 showed the largest deviation ($|\varepsilon_R| = 3.06$).

To visualize these predictions, Figure 4 compares the measured broadband spectra against two simulated baselines. The black lines represent the actual *in situ* measurements. The red lines represent the spectra generated using the parameters predicted by our 1D-CNN. For benchmarking, the blue dashed

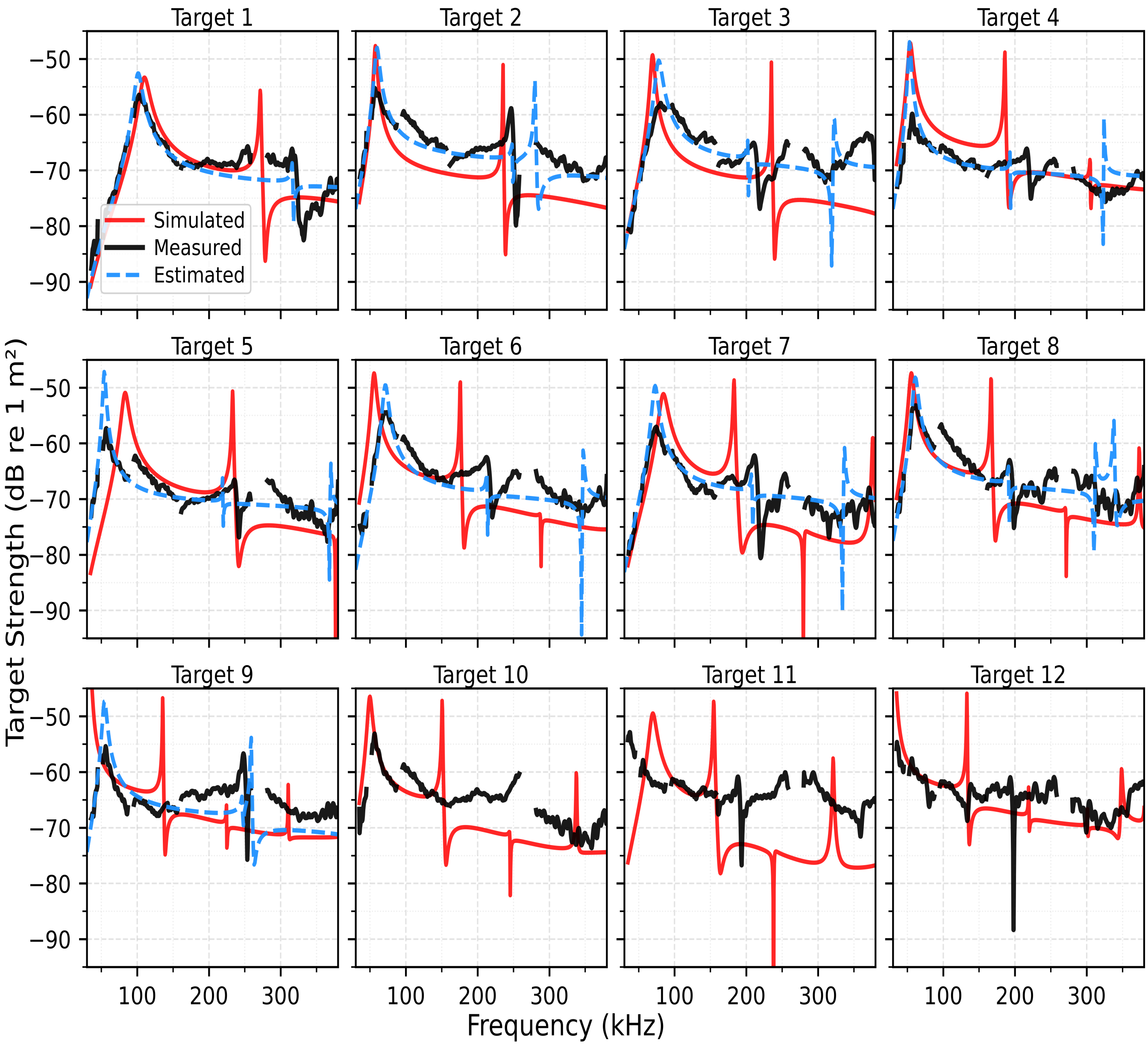


Figure 4. Comparison of measured *in situ* target strength spectra (black) against two simulated models. The red solid line represents spectra generated using the 1D-CNN predicted parameters. The blue dashed line represents spectra generated using the reference estimations from [35] with a fixed tilt angle of $\theta = 70°$.

lines show synthetic spectra generated using the analytical estimations from Khodabandeloo et al. [20], calculated with a fixed tilt angle of $\theta = 70°$. This comparison includes the nine labeled targets and three additional targets (10–12) for which reliable ground truth was not available due to incomplete resonance information.

For the majority of targets (1–8), the model closely predicted the frequency location of the first resonance peak. In several cases (e.g., Targets 2 and 4), the ML prediction aligns closely with the reference estimation (blue), confirming that the neural network has successfully learned to approximate the location of the first resonance peak. However, a systematic discrepancy is evident in the amplitude of the backscatter. As seen in Figure 4, the simulated spectra (red and blue) consistently exhibit sharper and higher-amplitude peaks than the measured *in situ* data (black). This “amplitude gap” indicates that while the model successfully learned the frequency-dependent features related to size, the physics-based training data lacked the damping factors present in real biological tissues. Consequently, the model predicts the correct location of the first resonance peak but the amplitude is exaggerated.

TABLE III. PERFORMANCE OF REGRESSION MODEL PREDICTIONS AGAINST *in situ* MEASUREMENTS. LABELS FOR ELONGATION $(\alpha+\beta)/2$ AND RADIUS $R_{ES}$ ARE TAKEN FROM KHODABANDELOO ET AL. [35]. ERRORS ARE ABSOLUTE DIFFERENCES BETWEEN PREDICTIONS AND LABELS.

| | Predictions | | | | | Labels | | Abs. Errors | |
|---|---|---|---|---|---|---|---|---|---|
| **Target** | $\boldsymbol{\theta}$ (°) | $\boldsymbol{a}$ (mm) | $\boldsymbol{b}$ (mm) | $\boldsymbol{\alpha} = \frac{a}{b}$ | $\boldsymbol{R}_{\text{eq}}$ (mm) | $(\boldsymbol{\alpha}+\boldsymbol{\beta})/2$ | $\boldsymbol{R}_{ES}$ (mm) | $\lvert\varepsilon_\alpha\rvert$ | $\lvert\varepsilon_R\rvert$ (mm) |
| 1 | 36.55 | 0.54 | 0.14 | 3.83 | 0.220 | 2.43 | 0.221 | 1.40 | 0.001 |
| 2 | 67.23 | 0.69 | 0.42 | 1.67 | 0.492 | 1.16 | 0.424 | 0.51 | 0.068 |
| 3 | 61.21 | 0.87 | 0.20 | 4.32 | 0.328 | 2.93 | 0.328 | 1.39 | 0.000 |
| 4 | 41.16 | 0.54 | 0.17 | 3.13 | 0.250 | 3.68 | 0.264 | 0.55 | 0.014 |
| 5 | 54.68 | 0.56 | 0.14 | 3.88 | 0.225 | 2.90 | 0.268 | 0.98 | 0.043 |
| 6 | 62.17 | 0.76 | 0.32 | 2.36 | 0.427 | 2.57 | 0.321 | 0.21 | 0.106 |
| 7 | 37.43 | 0.60 | 0.20 | 2.94 | 0.290 | 2.80 | 0.316 | 0.14 | 0.026 |
| 8 | 56.06 | 0.81 | 0.36 | 2.26 | 0.471 | 2.06 | 0.408 | 0.20 | 0.063 |
| 9 | 84.09 | 1.01 | 0.25 | 4.12 | 0.392 | 1.06 | 0.446 | 3.06 | 0.054 |
| 10 | 71.92 | 0.89 | 0.40 | 2.25 | 0.520 | - | - | - | - |
| 11 | 82.30 | 0.94 | 0.66 | 1.42 | 0.742 | - | - | - | - |
| 12 | 78.77 | 0.94 | 0.54 | 1.73 | 0.649 | - | - | - | - |

## IV. DISCUSSION

### A. Analysis of Model Performance

The tested machine learning models, including shallow learners (KNN, SVM, and RF) and a deep learner (1D-CNN), successfully classified gas- versus liquid-filled targets. This is primarily due to the distinct differences in the TS(f) responses between gas- and liquid-filled targets, as observed in Figure 1. In contrast, the models showed varying levels of performance in the regression task. Among them, the 1D-CNN demonstrated the highest accuracy in estimating the model parameters, evidenced by high $R^2$ scores for a, b, and $\theta$ (see Table II).

However, some outliers were observed in the predictions, where the parameters were not correctly estimated. This was especially the case for gas-filled targets in the prediction of angle $\theta$ and the computed aspect ratio $\alpha$. Further investigation of the outliers revealed that they mostly corresponded to small targets (i.e., small $\boldsymbol{R}_{\mathbf{eq}}$) (see [36] for more details). The TS frequency response of small gas-filled targets is known to be insensitive to shape and incident angle near the resonance frequency [37], explaining the model's difficulty in the 10–260 kHz frequency band. An example TS(f) response of such small targets is shown in the lower right panel of Figure 1. For deep learning approaches such as neural networks, large amounts of data are typically required; therefore, the use of optimized code for computing TS(f) was critical. Although the shallow learners achieved lower $R^2$ scores, it is important to note that less effort was devoted to tuning these models compared to the 1D-CNN. It is possible that more extensive hyperparameter optimization and validation could lead to improved performance for the shallow models.

Despite the strong baseline results, the application of the model to *in situ* measurements (Targets 1–12) revealed the complexities of the "reality gap" between idealized physics-based simulations and real-world acoustic environments.

### B. The Reality Gap: Frequency vs. Amplitude

The most significant finding from the *in situ* evaluation (Experiment 2) is the model's high accuracy in predicting the equivalent radius ($\boldsymbol{R}_{\mathbf{eq}}$), achieving an MAE of 0.042 mm. This success is directly linked to the model's ability to identify the frequency location of the first resonance peak (see Figure 4). This suggests that the model successfully learned size-dependent spectral features that are robust enough to transfer from idealized simulations to real-world measurements.

However, as seen in the spectral reconstructions, an "amplitude gap" persists where the model consistently overestimates Target Strength. This discrepancy likely stems from the physical limitations of the fluid-filled prolate spheroid model, which does not account for damping effects caused by, for example, surrounding tissue that absorbs part of the scattering due to viscous mechanisms that affect the peak amplitudes [38] of real marine organisms [37]. Furthermore, while the model identifies the first peak well, it struggles to precisely locate the second resonance peak in most targets (with the exception of Target 5). In contrast, the synthetic spectra generated from manual estimations (blue dashed lines) align more closely with these secondary features.

This observation extends to targets 10, 11, and 12, which lack established ground truth because they do not have a clear second resonance peak. For Target 10, the model accurately predicts the first resonance, and for Target 12, it captures the second peak location, yet both suffer from exaggerated amplitudes. Future iterations should improve accuracy by incorporating more sophisticated material properties into the scattering simulations. In addition, subsequent models should place greater emphasis on invariant features. Specifically, the relative ratio and location of the first and second resonance peaks can be identified independently of the absolute amplitude of the frequency response and are not affected by organism orientation [19].

### C. Scalability vs. Explainability

There is a clear operational advantage to using the 1D-CNN approach (red lines) over traditional manual resonance estimation (blue lines). Manual estimation is a labor-intensive process requiring significant human expertise and time per target, making it impractical for the large-scale, real-time datasets generated by modern broadband echosounders. Once trained, the ML model can provide near-instantaneous parameter estimations.

However, this efficiency introduces a trade-off in explainability. Manual estimation allows a researcher to justify a classification based on visible physical indicators (e.g., "the distance between these two peaks implies this elongation"). In contrast, deep learning architectures are often viewed as "black boxes" [39]. While the 1D-CNN is faster, the inability to easily verify the logic behind its predictions remains a challenge.

### D. Computational Feasibility

Finally, the 47.6-fold speedup achieved through NumPy vectorization was a critical enabler of this research. Deep learning architectures are data-intensive; without the ability to generate tens of thousands of spectra in a reasonable timeframe, the training of the hybrid 1D-CNN would have been unfeasible. This optimization was the result of a systematic evaluation of several techniques. Initial attempts to utilize GPU acceleration through libraries like CuPy [40] and PyTorch [41] provided negligible improvements, as profiling revealed that the overhead of frequent data transfers between the CPU and GPU created a significant performance bottleneck. Similarly, standard Python concurrency strategies yielded minor speed improvements; however, large-scale concurrency proved challenging due to several constraints. The Fortran-based subprocesses created Input/Output bottlenecks, while file-access contention and sequential dependency on externally generated files limited performance gains. Additionally, computational complexity increased with frequency, creating an imbalanced workload distribution. These challenges made further concurrency ineffective without extensive refactoring of the underlying code.

Vectorization with NumPy emerged as the most effective solution by replacing iterative Python loops with array-based operations optimized in C. This approach allowed for the simultaneous calculation of entire frequency bands while maintaining identical numerical precision to the original code. This demonstrates that in simulation-driven ML, technical optimization of the physics engine is just as vital as the neural network architecture itself. While the current implementation was sufficient for this study, future iterations requiring more complex, inhomogeneous geometries may benefit from just-in-time (JIT) compilation or low-level C++ implementations to further bridge the gap between computational cost and model complexity. These methods could offer further speedups beyond standard vectorization, facilitating the modeling of even more complex marine organisms without the prohibitive computational costs.

## V. Conclusion and Future Work

This study demonstrates that machine learning, specifically a 1D-CNN trained exclusively on physics-based simulated data, can accurately classify and estimate geometric properties of acoustic targets. We demonstrated that the hybrid 1D-CNN outperforms traditional machine learning methods such as KNN, SVM, and RF. While all models achieved perfect classification of gas- and liquid-filled targets, the 1D-CNN was notably more accurate in the regression task of estimating the targets' geometric properties, including semi-major/minor axes (a, b) and incident angle ($\theta$).

The feasibility of this approach was facilitated by the computational optimization of the scattering model through NumPy vectorization. By achieving a 47.6-fold speed increase, it was possible to generate the massive synthetic datasets required for deep learning within a practical timeframe.

Outlier analysis revealed specific challenges, particularly in estimating geometric parameters of small, gas-filled targets, with simpler spectral features lacking distinct resonance peaks. Addressing these limitations by expanding frequency ranges or incorporating additional acoustic parameters in future research could enhance performance further.

Validation on real-world *in situ* measurements showed that the model can generalize to noisy environments, particularly for size estimation, where it achieved a mean absolute error for the equivalent radius of only 0.042 mm. Addressing current limitations, such as the overestimation of peak amplitudes by incorporating damping effects and more realistic scattering effects in future research could enhance performance further.

We conclude that this simulation-driven approach is a powerful and viable strategy for overcoming data scarcity in marine acoustics. It represents a promising step toward the automated, non-invasive classification of marine organisms, with potential application in real-time classification of mesopelagic organisms during acoustic surveys, supporting biomass estimation and reducing the need for manual analysis.

## Acknowledgment

S. M. acknowledges the use of the "Birget" high-performance computing cluster, provided by the Department of Informatics, University of Bergen, which was essential for this research. B. K. and K. M. gratefully acknowledge funding from the Center for Research-based Innovation in Marine Acoustic Abundance Estimation and Backscatter Classification (CRIMAC, Project No. 309512; www.crimac.no). This work is based on the master's thesis of S. M. conducted at the University of Bergen.

## References

[1] S. A. B. Marx, B. Khodabandeloo, and K. Malde, "Machine learning using physics-based backscattering modeling for acoustic identification of mesopelagic organisms", in *Proceedings of COCE 2025: The Second International Conference on Technologies for Marine and Coastal Ecosystems*, Barcelona, Spain: IARIA, Oct. 2025, pp. 32–37, ISBN: 978-1-68558-329-3.

[2] C. F. Greenlaw, "Acoustical estimation of zoo-plankton populations", *Limnology and Oceanography*, vol. 24, pp. 226–242, 1979.

[3] A. C. Lavery et al., “Determining dominant scatterers of sound in mixed zooplankton populations”, *J. Acoust. Soc. Am.*, vol. 122, pp. 3304–3326, 2007.

[4] R. J. Kloser, T. E. Ryan, J. W. Young, and M. E. Lewis, “Acoustic observations of micronekton fish on the scale of an ocean basin: potential and challenges”, *ICES J. Mar. Sci.*, vol. 66, pp. 998–1006, 2009. DOI: 10.1093/icesjms/fsp077

[5] R. Hickling, “Analysis of echoes from a solid elastic sphere in water”, *J. Acoust. Soc. Am.*, vol. 34, pp. 1582–1592, 1962. DOI: 10.1121/1.1909055

[6] J. J. Faran, “Sound scattering by solid cylinders and spheres”, *J. Acoust. Soc. Am.*, vol. 23, pp. 405–417, 1951. DOI: 10.1121/1.1906780

[7] D. Chu and T. K. Stanton, “Application of pulse compression techniques to broadband acoustic scattering by live individual zooplankton”, *J. Acoust. Soc. Am.*, vol. 104, pp. 39–55, 1998. DOI: 10.1121/1.424056

[8] M. E. Zakharia, “Wide band fisheries sounder; From individual echoes analysis to classification of schools at sea”, *J. Acoust. Soc. Am.*, vol. 103, pp. 3068–3068, 1998. DOI: 10.1121/1.422849

[9] J. E. Ehrenberg and T. C. Torkelson, “FM slide (chirp) signals: a technique for significantly improving the signal-to-noise performance in hydroacoustics assessment systems”, *Fisheries Research*, vol. 47, pp. 193–199, 2000.

[10] A. C. Lavery, T. K. Stanton, D. E. McGehee, and D. Chu, “Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton”, *J. Acoust. Soc. Am.*, vol. 111, pp. 1197–1210, 2002. DOI: 10.1121/1.1433813

[11] J. M. Jech et al., “Comparisons among ten models of acoustic backscattering used in aquatic ecosystem research”, *J. Acoust. Soc. Am.*, vol. 138, pp. 3742–3764, 2015. DOI: 10.1121/1.4937607

[12] L. Alzubaidi et al., “A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications”, *Journal of Big Data*, vol. 10, p. 46, 2023, Art. 46. DOI: 10.1186/s40537-023-00727-2

[13] R. Adelman, N. A. Gumerov, and R. Duraiswami, “Semi-analytical computation of acoustic scattering by spheroids and disks”, *The Journal of the Acoustical Society of America*, vol. 136, no. 6, EL405–EL410, Nov. 2014, ISSN: 0001-4966. DOI: 10.1121/1.4901318 eprint: https://pubs.aip.org/asa/jasa/article-pdf/136/6/EL405/13131806/el405_1_online.pdf.

[14] H. Medwin, *Sounds in the sea: From ocean acoustics to acoustical oceanography*. Cambridge University Press, 2005.

[15] R. H. Love, “Resonant acoustic scattering by swimbladder-bearing fisha)”, *The Journal of the Acoustical Society of America*, vol. 64, no. 2, pp. 571–580, Aug. 1978, ISSN: 0001-4966. DOI: 10.1121/1.382009 eprint: https://pubs.aip.org/asa/jasa/article-pdf/64/2/571/11724662/571_1_online.pdf.

[16] M. D. Agersted, B. Khodabandeloo, Y. Liu, W. Melle, and T. A. Klevjer, “Application of an unsupervised clustering algorithm on in situ broadband acoustic data to identify different mesopelagic target types”, *ICES Journal of Marine Science*, vol. 78, no. 8, pp. 2907–2921, Aug. 2021, ISSN: 1054-3139. DOI: 10.1093/icesjms/fsab167 eprint: https://academic.oup.com/icesjms/article-pdf/78/8/2907/41764791/fsab167.pdf.

[17] R. Kubilius, G. J. Macaulay, and E. Ona, “Remote sizing of fish-like targets using broadband acoustics”, *Fisheries Research*, vol. 228, p. 105 568, 2020, Art. 105568. DOI: 10.1016/j.fishres.2020.105568

[18] S. Kaartvedt, A. Staby, and D. Aksnes, “Efficient trawl avoidance by mesopelagic fishes causes large underestimation of their biomass”, *Marine Ecology Progress Series*, vol. 456, pp. 1–6, Jun. 2012. DOI: 10.3354/meps09785

[19] B. Khodabandeloo, Y. Heggelund, B. Ystad, S. A. B. Marx, and G. Pedersen, “High-precision model and open-source software for acoustic backscattering by liquid- and gas-filled prolate spheroids across a wide frequency range and incident angles: Implications for fisheries acoustics”, *Journal of Sound and Vibration*, vol. 616, p. 119 227, 2025, ISSN: 0022-460X. DOI: https://doi.org/10.1016/j.jsv.2025.119227

[20] B. Khodabandeloo et al., “Mesopelagic fish gas bladder elongation, as estimated from wideband acoustic backscattering measurements”, *The Journal of the Acoustical Society of America*, vol. 151, no. 6, pp. 4073–4085, Jun. 2022, ISSN: 0001-4966. DOI: 10.1121/10.0011742 eprint: https://pubs.aip.org/asa/jasa/article-pdf/151/6/4073/16529753/4073_1_online.pdf.

[21] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance”, *Applied Soft Computing*, vol. 97, p. 105 524, 2020, Art. 105524. DOI: 10.1016/j.asoc.2019.105524

[22] S. Kiranyaz et al., “1D convolutional neural networks and applications: A survey”, *Mechanical Systems and Signal Processing*, vol. 151, p. 107 398, 2021, Art. 107398. DOI: 10.1016/j.ymssp.2020.107398

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition”, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90

[24] J. Hu, L. Shen, and G. Sun, “Squeeze-and-excitation networks”, in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141. DOI: 10.1109/CVPR.2018.00745

[25] S. Ioffe and C. Szegedy, “Batch normalization: accelerating deep network training by reducing internal covariate shift”, in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, 2015, pp. 448–456.

[26] E. Fix and J. L. Hodges, “Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties”, *International Statistical Review / Revue Internationale de Statistique*, vol. 57, pp. 238–247, 1989. DOI: 10.2307/1403797

[27] E. Cotter, C. Bassett, and A. Lavery, “Classification of broadband target spectra in the mesopelagic using physics-informed machine learning”, *J. Acoust. Soc. Am.*, vol. 149, pp. 3889–3901, 2021. DOI: 10.1121/10.0005114

[28] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines”, *IEEE Intelligent Systems and Their Applications*, vol. 13, pp. 18–28, 1998. DOI: 10.1109/5254.708428

[29] H. Yang et al., “Underwater acoustic target recognition using SVM ensemble via weighted sample and feature selection”, in *Proc of the 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2016, pp. 522–527. DOI: 10.1109/IBCAST.2016.7429928

[30] L. Breiman, “Random Forests”, *Machine Learning*, vol. 45, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324

[31] S. M. Gugele et al., “Differentiation of two swim bladdered fish species using next generation wideband hydroacoustics”, *Scientific Reports*, vol. 11, p. 10 520, 2021, Art. 10520. DOI: 10.1038/s41598-021-89941-7

[32] B. Khodabandeloo, M. D. Agersted, T. Klevjer, G. J. Macaulay, and W. Melle, “Estimating target strength and physical characteristics of gas-bearing mesopelagic fish from wideband in situ echoes using a viscous-elastic scattering model”, *The Journal of the Acoustical Society of America*, vol. 149, no. 1, pp. 673–691, Jan. 2021, ISSN: 0001-4966. DOI: 10.1121/10.0003341 eprint: https://pubs.aip.org/asa/jasa/article-pdf/149/1/673/19745391/673_1_online.pdf.

[33] Y. A. Cengel and M. A. Boles, *Thermodynamics: An Engineering Approach*, 4th ed. Singapore: McGraw-Hill, 2002.

[34] P. J. Huber, "Robust Estimation of a Location Parameter", *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964. DOI: 10.1214/aoms/1177703732

[35] B. Khodabandeloo et al., "Mesopelagic fish gas bladder elongation, as estimated from wideband acoustic backscattering measurements", *J. Acoust. Soc. Am.*, vol. 151, pp. 4073–4085, 2022. DOI: 10.1121/10.0011742

[36] S. A. B. Marx, "Learning acoustic target classification from simulation", Master's Thesis, University of Bergen, 2025.

[37] B. Khodabandeloo, M. D. Agersted, T. A. Klevjer, G. Pedersen, and W. Melle, "Mesopelagic flesh shear viscosity estimation from in situ broadband backscattering measurements by a viscous–elastic model inversion", *ICES Journal of Marine Science*, vol. 78, no. 9, pp. 3147–3161, 2021. DOI: 10.1093/icesjms/fsab183

[38] C. Feuillade and R. W. Nero, "A viscous-elastic swimbladder model for describing enhanced-frequency resonance scattering from fish", *The Journal of the Acoustical Society of America*, vol. 103, no. 6, pp. 3245–3255, Jun. 1998, ISSN: 0001-4966. DOI: 10.1121/1.423076 eprint: https://pubs.aip.org/asa/jasa/article-pdf/103/6/3245/8084717/3245_1_online.pdf.

[39] R. Guidotti et al., "A survey of methods for explaining black box models", *ACM Comput. Surv.*, vol. 51, no. 5, 2018, ISSN: 0360-0300. DOI: 10.1145/3236009

[40] R. Okuta, Y. Unno, D. Nishino, S. Hido, and C. Loomis, "Cupy: A numpy-compatible library for NVIDIA GPU calculations", in *Proceedings of Workshop on Machine Learning Systems (LearningSys) in The Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*, 2017.

[41] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library", in *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., 2019, pp. 8024–8035.

[42] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. DOI: 10.1016/j.jcp.2018.10.045