Machine Learning Using Physics-Based Backscattering Modeling for Acoustic Identification of Mesopelagic Organisms

Sander Andre Berg Marx[1,*] , Babak Khodabandeloo[2,†], Ketil Malde[1,2,‡]
[1]Department of Informatics, University of Bergen, Bergen, Norway
[2]Acoustics and Observation Methodologies, Institute of Marine Research (IMR), Bergen, Norway
[*] email: sanderamarx@hotmail.com
[†] email: babak.khodabandeloo@hi.no
[‡] email: 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 Forrest (RF). Results show that while all models achieved perfect classification accuracy, the hybrid 1D-CNN clearly outperformed the others in parameter estimation. This demonstrates that simulation-driven machine learning can help overcome data scarcity and enable automated acoustic identification of mesopelagic organisms.**

Keywords-Acoustic target classification; machine learning; Convolutional neural network; prolate spheroid backscattering modeling.

I. INTRODUCTION

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 [1][2][3]. 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 [4][5]. 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 [6]. Broadband fisheries echosounders transmit frequency-modulated pulses, offering two main advantages [7][8]: (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 [9]. 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 [10]. However, their high computational cost can limit their practical use. 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 key characteristics such as size, elongation, and orientation. For example, spherical models have long been used in fisheries acoustics due to their simplicity [11][12]. 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. However, large, accurately labeled datasets are needed to develop robust machine learning models [13]. Acquiring such datasets is challenging and expensive in real-world marine environments [14].

While real acoustic datasets exist, they rarely provide reliable ground truth. To overcome this limitation, we use a physics-based simulation model to generate large synthetic datasets for training. On this basis, a novel hybrid one-Dimensional Convolutional Neural Network (1D-CNN) is proposed for simultaneous classification (gas-filled vs liquid) 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 remainder of the paper is organized as follows. Section II describes the methods for data generation and machine learning model training. Section III reports the results, and Section IV provides a discussion of the findings. Section V concludes the paper and outlines directions for future work.

II. METHOD

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 [15]. 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. 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 R_{eq}) and elongation (or aspect ratio α) instead of directly using semi major and minor axes, a and b, respectively. They are related by the following equations:

$$b = R_{eq}/(\alpha^{1/3}), \quad (1)$$

$$a = \alpha \times b\,. \quad (2)$$

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 (*Req*)	0.1–5 mm	5–20 mm
Aspect ratio (α)	1.05–8.0	1.05–8.0
Spheroid Density (ρ_s)	2–80 kg/m^3	1.01–1.07 × (ρ_w) kg/m^3
Water Density (ρ_w)	1027 kg/m^3	1027 kg/m^3
Spheroid Sound Speed (c_s)	343 m/s	1.01-1.07 × (c_w) m/s
Water Sound Speed (c_w)	1500 m/s	1500 m/s
Incident angle (θ)	0.01–90°	0.01–90°
Frequency range	10–260 kHz (0.5 kHz steps)	

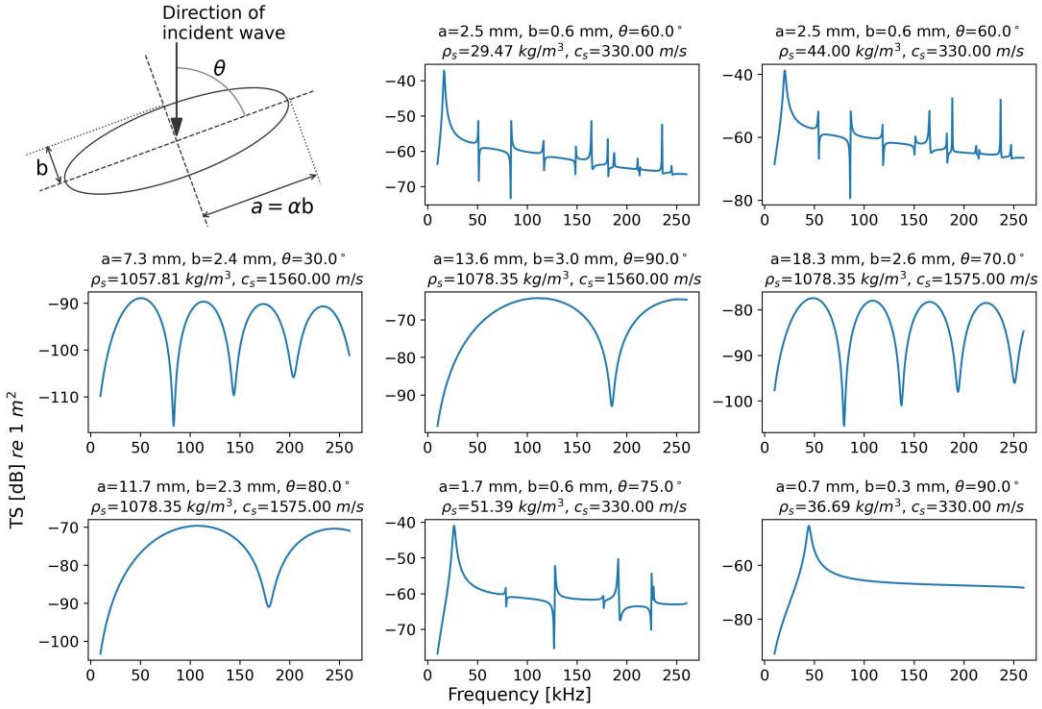

Figure 1. Examples of TS(f) for prolate spheroids with parameters within the ranges in Table I. An illustration of a prolate spheroid with semi-major and semi-minor axes, a and b respectively, along with the incident angle θ, is shown in the upper left corner.

B. *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 (θ), 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 [16]. 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 (θ, 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.

C. *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, like acoustic frequency spectra [17].

The network architecture, illustrated in Figure 2, was designed to process sequential TS data. The architecture incorporates advanced deep learning components, specifically residual blocks [18] and Squeeze-and-Excitation (SE) attention mechanisms [20], to improve feature extraction and training stability. The TS spectra is 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 [19] and ReLU activations. Also integrated in these residual blocks are Squeeze and Excitation (SE) attention mechanisms that perform channel-wise feature weighting by compressing information through global average pooling and then learning channel relationships via two fully connected layers. 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 classification of target type, one for regression of the incident angle (θ) 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) [21]: 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. [22].
- Support Vector Machine (SVM) [23]: Seeking an optimal hyperplane to separate classes or predict continuous values, utilized by Yang et al. [24] for underwater target recognition.
- Random Forest (RF) [25]: An ensemble method constructing multiple decision trees on random subsets of data and features, aggregating their predictions, as employed by Gugele et al. [26].

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

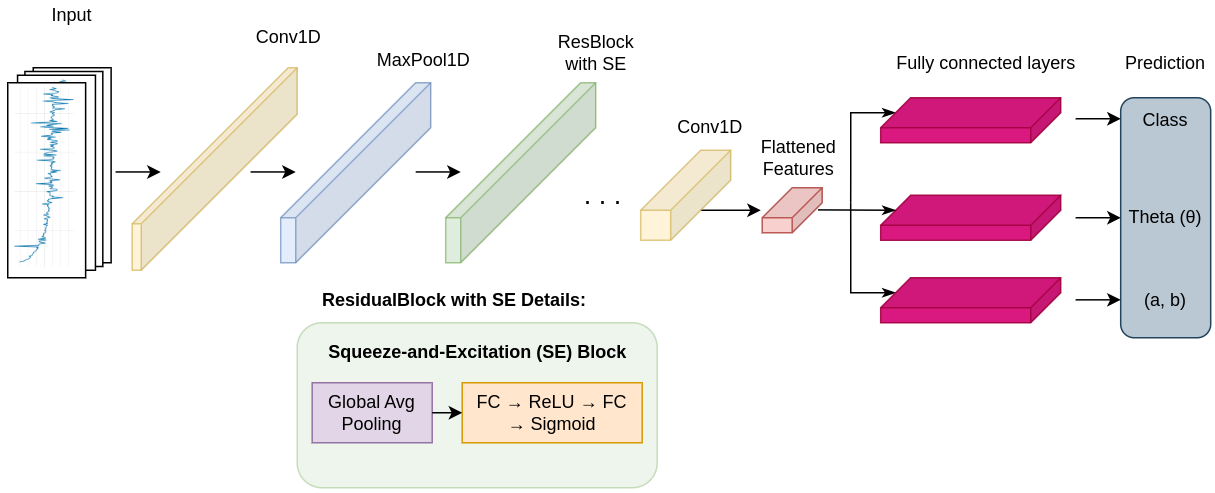

Figure 2. Hybrid one-dimensional convolutional neural network architecture for acoustic target classification.

D. *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 (Λ) is a weighted sum of four components as shown in (3):

$$\Lambda = \omega_c \cdot \Lambda_c + \omega_r \cdot (\Lambda_\theta + \Lambda_{ab}) + \omega_{cons} \cdot \Lambda_{cons}. \quad (3)$$

where Λ_c is cross-entropy loss for the binary classification task. For regression, Huber loss [27] was used for the angle (Λ_θ) and Mean Squared Error (MSE) for the size parameters (Λ_{ab}). To enforce the geometric relationship between the semi-major and minor axes, a consistency loss ($\Lambda_{cons.}$), 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 θ, a, b, and the derived α. 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. *Performance Comparison Across Models*

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 θ	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**

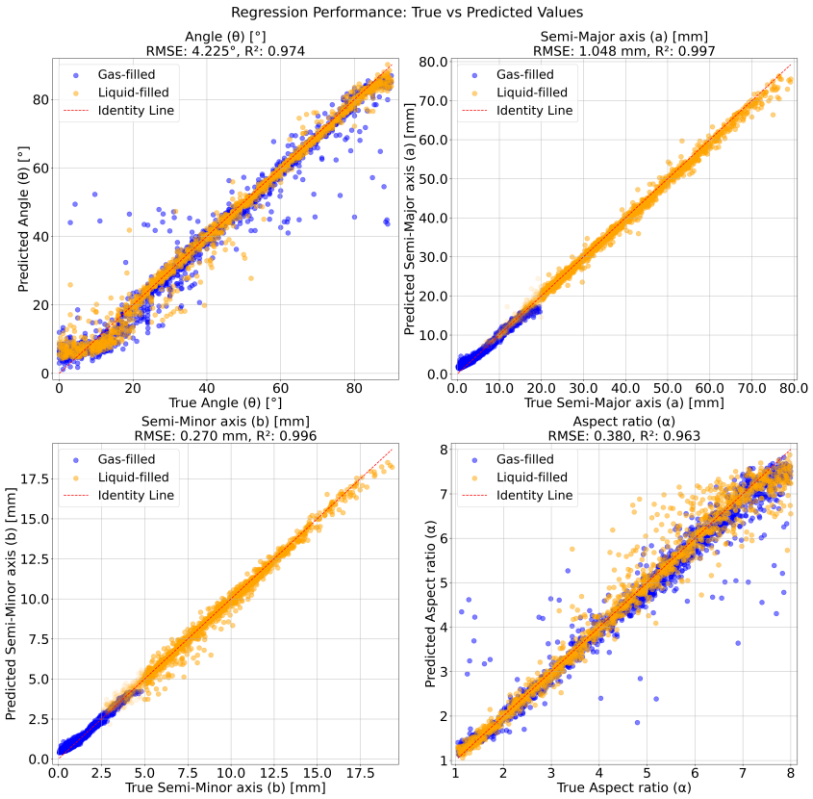

Figure 3. Regression performance of the hybrid 1D-CNN model on test data showing true values against predicted values for (top-left) angle (θ), (top-right) semi-major axis (a), (bottom-left) semi-minor axis (b), and (bottom-right) dervided aspect ratio (α = a/b). Blue points represent gas-filled targets,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

B. 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 θ 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 ($\alpha > 4$).

C. Analysis of prediction outliers

Outlier analysis revealed consistent patterns for both incident angle (θ) and aspect ratio (α).

For incident angle (θ) 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 (α) 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 [28], a feature these outlier spectra lacked.

IV. DISCUSSION

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 θ (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 θ and the computed aspect ratio α. Further investigation of the outliers revealed that they mostly corresponded to small targets (i.e. small R_{eq}) (see [29] 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 [30], 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.

V. CONCLUSION AND FUTURE WORK

This study demonstrates that using machine learning, more particularly 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 (θ).

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. Although this study focused on simulated data, preliminary tests on a small real dataset [29] showed encouraging results, indicating potential for application to in situ measurements.

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] C. F. Greenlaw, "Acoustical estimation of zoo-plankton populations", Limnol. Oceanogr., vol. 24, pp. 226–242, 1979.

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

[3] 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, https://doi.org/10.1093/icesjms/fsp077.

[4] R. Hickling, "Analysis of echoes from a solid elastic sphere in water", J. Acoust. Soc. Am., vol. 34, pp. 1582–1592, 1962, https://doi.org/10.1121/1.1909055.

[5] J.J. Faran, "Sound scattering by solid cylinders and spheres", J. Acoust. Soc. Am., vol. 23, pp. 405–417, 1951, https://doi.org/10.1121/1.1906780.

[6] 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, https://doi.org/10.1121/1.424056.

[7] 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, https://doi.org/10.1121/1.422849.

[8] 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.

[9] 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 https://doi.org/10.1121/1.1433813.

[10] 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, https://doi.org/10.1121/1.4937607.

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

[12] R. H. Love, "Resonant acoustic scattering by swimbladder-bearing fish", J. Acoust. Soc. Am., vol. 64, pp. 571–580, 1978.

[13] 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, art. 46, 2023, https://doi.org/10.1186/s40537-023-00727-2.

[14] R. Kubilius, G. J. Macaulay, and E. Ona, "Remote sizing of fish-like targets using broadband acoustics", Fisheries Research vol. 228, art. 105568, 2020, https://doi.org/10.1016/j.fishres.2020.105568.

[15] 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, art. 119227, 2025, https://doi.org/10.1016/j.jsv.2025.119227.

[16] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance", Applied Soft Computing vol. 97, art. 105524, https://doi.org/10.1016/j.asoc.2019.105524.

[17] S. Kiranyaz et al., "1D convolutional neural networks and applications: A survey", Mechanical Systems and Signal Processing, vol. 151, art. 107398, 2021, https://doi.org/10.1016/j.ymssp.2020.107398.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016, https://doi.org/10.1109/CVPR.2016.90.

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

[20] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks", IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141, 2018, https://doi.org/10.1109/CVPR.2018.00745.

[21] 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 https://doi.org/10.2307/1403797.

[22] E. Cotter, C. Bassett, and A. Lavery, "Classification of broadband target spectra in the mesopelagic using physics-informed machine learning", The Journal of the Acoustical Society of America, vol. 149, pp. 3889–3901, 2021, https://doi.org/10.1121/10.0005114.

[23] 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, https://doi.org/10.1109/5254.708428.

[24] H. Yang et al., "Underwater acoustic target recognition using SVM ensemble via weighted sample and feature selection", 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST), pp. 522–527, 2016, https://doi.org/10.1109/IBCAST.2016.7429928.

[25] L. Breiman, "Random Forests", Machine Learning, vol. 45, pp. 5–32, 2001, https://doi.org/10.1023/A:1010933404324.

[26] S. M. Gugele et al., "Differentiation of two swim bladdered fish species using next generation wideband hydroacoustics", Sci Rep, vol. 11, art. 10520, 2021, https://doi.org/10.1038/s41598-021-89941-7.

[27] P. J. Huber, "Robust Estimation of a Location Parameter," Ann. Math. Statist. vol. 35, pp. 73 - 101, 1964, https://doi.org/10.1214/aoms/1177703732

[28] 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, https://doi.org/10.1121/10.0011742.

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

[30] 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, https://doi.org/10.1093/icesjms/fsab183.