# EVALUATION OF A MACHINE LEARNING APPROACH FOR UNDERWATER TARGET CLASSIFICATION WITH LOW-FREQUENCY ACTIVE SONAR ROBUST TO ENVIRONMENT DIFFERENCES

HJ Kuijf TNO, the Hague, the Netherlands
AM van Heteren TNO, the Hague, the Netherlands
VO Oppeneer TNO, the Hague, the Netherlands
MC van Leeuwen TNO, the Hague, the Netherlands
R van Vossen TNO, the Hague, the Netherlands

## 1 INTRODUCTION

Low-Frequency Active Sonar (LFAS) systems are the method of choice for long-range underwater surveillance and detection of targets. To extract relevant sonar contacts, the raw LFAS data undergoes several signal processing steps. These include processes such as beamforming, matched filtering, detrending of the signal as a function of range, and additional analyses to extract candidate sonar contacts [1]. Each candidate contact is classified as either background/clutter or a potential target.

Feature-based classification and machine learning or artificial intelligence technology can assist with (semi)automatic classification of sonar contacts. Such systems can be trained with a dataset of measured target data [2] [3] or employed as anomaly detection trained only in a given environment [4]. In both situations, systems heavily rely on the provided recorded example data. Deploying a trained system in a new underwater environment will likely lead to degraded performance. Unfortunately, new LFAS target data is expensive to obtain, environmentally dependent, and scarce. Simulating sonar data in different environments is a useful approach to generate new data [5]. Nevertheless, the key question remains: what is the optimal simulation and training strategy given the environmental dependency?

In this work, three different approaches of applying trained machine learning systems in a new underwater environment are studied. Various underwater environments will be simulated and include clutter (simulated solid sphere made of rock) and targets (simulated air-filled steel spheres). Our approaches include: (1) a zero-shot/generalization approach that is evaluated in an environment not present in the training data, (2) a few-shot approach in which the zero-shot classifier is fine-tuned with a subset of in-situ sonar data, and (3) an environment-specific approach that is only trained on a subset of in-situ sonar data and evaluated in that same environment. A number of machine learning systems will be investigated, including a support-vector machine (SVM), a fully-connected neural network (FCN), and a convolutional neural network (CNN).

## 2 METHODS

### 2.1 Data simulation

The clutter and target data are simulated by combining artificial background data and a target echo time series of this target at pre-specified ranges, bearings, and depths in this environment. The data simulation can be divided into two parts: (i) the background simulation, and (ii) the target echo simulation and injection.

### 2.1.1 Background simulation

To determine the effect of the environment, four general environmental descriptions were considered, partly based on the Weston Memorial workshop [6] [7]:

- Weston 1: Shallow water, Iso sound speed profile
- Weston 2: Shallow water, winter profile (linear profile)
- Weston 3: Shallow water, summer profile (weak surface duct)
- Weston 7: Deep water, Munk profile

The sediment type is very important for the background, because it changes how much energy remains inside the water column and how much energy is lost to the sediment. In the simulations, the following sediment types are considered [8]: course clay, medium silt, medium sand, and very coarse sand.

The depths of the environments are varied as well. The water depths for the shallow water environments are 50, 100, 150, 200, and 250 m. For the deep-water scenario, the water depths are 500, 1000, 1500, 2000, and 2500 m.

The noise in the scenarios are fully wind-generated, the sea state is assumed to be 3, and there is no shipping noise. The reverberation in the scenario follows Lambert's rule, with $\mu$=-27 dB [8, p. 500]. For the scenario, a transducer is chosen to be at 25 m depth, which can both transmit and receive. The transducer is a horizontal line array, pointing north. The transducer sends out an LFM waveform.

The TNO sonar performance model ARCANE [9] is used to compute the propagation, the noise, and the reverberation of this environment.

### 2.1.2 Simulation and injection of targets echoes

Before anything can be done with target injections, first a set of targets must be chosen and their locations (range, bearing, depth) have to be determined. Two types of objects are chosen: a solid sphere made of granite as a proxy for clutter, and an air-filled sphere of steel as a proxy for a target. For these objects, analytical formulas for a sphere in an infinite medium are used [10].

The simulation of target echoes requires to combine target scattering and propagation. This is done using a ray tracer, following [8, p. 615]. The echo level as a function of time is computed by tracing rays from the source to the target location, combined with the corresponding angle dependent target strength, and finally combined again with the propagation from the target to the receiver. As mentioned before, the transducer is a horizontal line array and has a beampattern. This beampattern is also applied to the received signal. Since this process is done in the frequency domain, the final step in getting the time signal is done by doing an inverse Fourier transform. The time signal then has to be injected into the precomputed background. This is done by converting the time axis of the signal to a range axis (by multiplying by the speed of sound) and fusing the target with the background. From this point on, the backgrounds with the injected targets can be put through the further processing chain.

## 2.2 Snippet processing chain

The data simulation produces a range-bearing matrix (or 'ping') with size 360×102400, corresponding to a bearing of -180° to 179° and a range of 15 km. An example can be seen in Figure 1. The data is processed using similar steps as described in [2]. The beamformed and matched filtered data is detrended and provided as input to a detector. The detector gives a number of range-bearing detected points per ping as output. For each detection point, a snippet matrix of 17×1053, centred around the detection point, was extracted. The snippet matrices contain normalized values in dB. Some examples are shown in Figure 1.

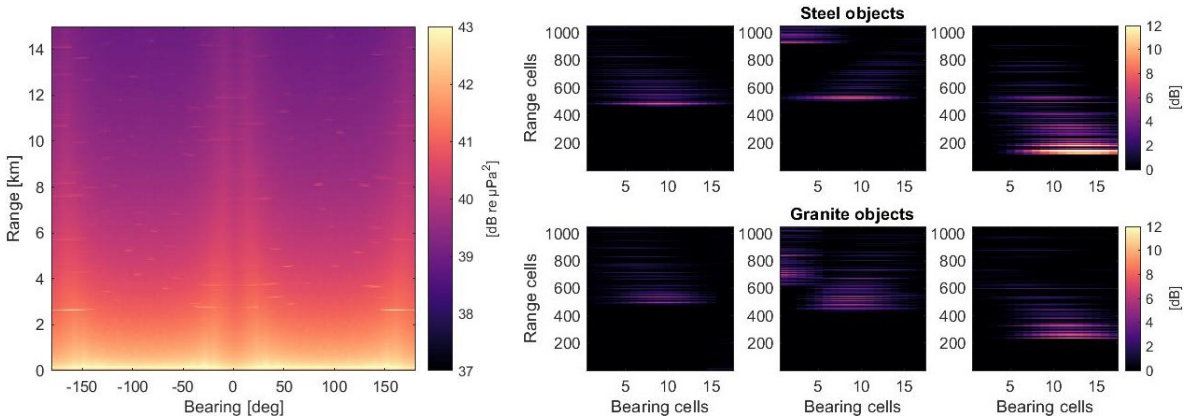


*Figure 1 Left: typical example of a range-bearing matrix. Right: examples of snippets corresponding to target echoes from air-filled spheres (top row) and from clutter (granite spheres; bottom row).*

### 2.3 Machine learning approaches

Three machine learning approaches were implemented: a support vector machine (SVM), a fully-connected neural network (FCN), and a convolutional neural network (CNN). These approaches will classify the snippets as either background/clutter or a target.

#### 2.3.1 Support vector machine (SVM)

A SVM was implemented using SGDclassifier with the hinge loss function from the Scikit-learn Python package [11]. The SGDclassifier with hinge loss was used instead of LinearSVC, because it scaled better with the number of data points. Candidate sonar contacts were flattened into a 1D array and concatenated with the range and bearing of the contacts, before providing them to the SVM. The number of training epochs was 250. For the zero-shot approach, the batch size was 10000 candidate sonar contacts and for the other two approaches, the batch size was 120 candidate sonar contacts.

#### 2.3.2 Fully-connected neural network (FCN)

A fully-connected neural network was implemented based on PyTorch Lightning [12] [13]. Candidate sonar contacts were flattened into a 1D array before providing them to the FCN. The FCN consisted of one input layer, a hidden layer of 8 nodes, a second hidden layer of 4 nodes, and an output layer. The range and bearing value of the contact was added to the hidden layer. The loss function was binary cross entropy, the optimizer was Adam and the learning rate was initialized at 0.0005. For the zero-shot approach, the batch size was 10000 candidate sonar contacts and the FCN was trained for 2000 epochs. For the other two approaches, the batch size was 120 candidate sonar contacts. The number of epochs for the few-shot approach was 150 and for the environment-specific approach 1500.

#### 2.3.3 Convolutional neural network

A convolutional neural network was implemented based on PyTorch Lightning [12] [13]. The CNN consisted of two convolutional layers followed by two fully connected layers. To account for the anisotropic pixel sizes in the range-bearing snippets, the first convolutional layer used eight kernels of size [2×17] with a stride of [1×1], followed by a flattening layer and an average pooling layer of size 2. This way, the information in the bearing direction of the snippet was extracted. The second, one-dimensional, convolutional layer used sixteen kernels of size 100 and stride 2, followed by an average pooling layer of size 4. The first fully connected layer consisted of 864 nodes and the second of 8 nodes. The range and bearing value of the contact was added to the second layer. The loss function was binary cross entropy, the optimizer was Adam and the learning rate was initialized at 0.001. For the zero-shot approach, the batch size was 10000 candidate sonar contacts, and the CNN was trained

for 2000 epochs. For the other two approaches, the batch size was 120 candidate sonar contacts. The number of epochs for the few-shot approach was 150 and for the environment-specific approach 1500.

### 2.4 Experiments and evaluation

#### 2.4.1 Simulation overview

As explained in Section 2.1.1, a large number of unique environments was considered. Per unique environment, 12 'pings' were simulated. Out of these pings, six included 100 steel spheres as a proxy for a target and the other six included 100 granite spheres as a proxy for clutter. The steel and granite spheres had a diameter of 10 m. The steel sphere was filled with air, has a thickness of 0.2 m, and was placed at a random position in the water column. The granite sphere was solid and was always placed at 1 m from the sea bottom. The objects were placed at random positions at a range between 2 and 15 km from the source, in random directions. Both source and receiver were placed at a depth of 25 m.

#### 2.4.2 Experiments

Three approaches will be evaluated with the simulated datasets:

1. A zero-shot / generalization approach, where the performance of the machine learning systems is evaluated in a dataset that is not present in the training data.
2. A few-shot approach, where the machine learning system from the zero-shot approach is fine-tuned with a limited subset of in-situ sonar data from the environment it will be evaluated in.
3. An environment-specific approach, that is trained with a limited subset of in-situ sonar data and evaluated in that exact same environment.

Each environment $E_{PSD}$ has three variables that can change: $P$ being the Weston sound speed profile (1, 2, 3, or 7), $S$ being the sediment type (course clay, medium silt, medium sand, and very coarse sand), and $D$ being the depth. See Section 2.1.1 for details on these variables. The data of every environment $E_{PSD}$ is split into a subset for training (60% of the data) and a subset for test (40% of the data).

All machine learning systems will be evaluated in a cross-validation setting, where the system is trained on a selection of the training data subsets and evaluated on the relevant test data subsets. In each cross-validation setting, one $PSD$ parameter will be varied and the others will be fixed. For example for the Weston sound speed profile $P$:

1. Zero-shot / generalization: the machine learning systems are trained with the training subsets from $E_{\{1,2,3\}\{all\}\{all\}}$ and evaluated on the test subsets of $E_{\{7\}\{all\}\{all\}}$. This is then repeated for the other Weston sound speed profiles $p \in P$.
2. Few-shot: similarly, the machine learning systems are evaluated in a cross-validation setting. The systems are trained with the training subsets from $E_{\{1,2,3\}\{all\}\{all\}}$ (like the zero-shot), then fine-tuned with the training subset from $E_{\{7\}\{s\}\{d\}}$, and evaluated on the test subset from $E_{\{7\}\{s\}\{d\}}$ (separately for every $s \in S$ and $d \in D$). This is then repeated for the other Weston sound speed profiles $p \in P$.
3. Environment-specific: : similarly, the machine learning systems are evaluated in a cross-validation setting. The systems are trained with the training subsets from $E_{\{7\}\{s\}\{d\}}$ and evaluated on the test subset from the same environment $E_{\{7\}\{s\}\{d\}}$ (separately for every $s \in S$ and $d \in D$). This is then repeated for the other Weston sound speed profiles $p \in P$.

This approach is repeated for the sediment types ($S$ is varied, while considering all $P$ and $D$) and the depths ($D$ is varied, while considering all $P$ and $S$). Receiver operating characteristic (ROC) curves will be generated for each approach and the area-under-the-curve (AUC) will be computed.

# 3 RESULTS

The snippet processing chain was applied to all simulated datasets. Because the signal-to-noise ratio can be low for contacts at larger range, not all injected targets were presented as a snippet. For evaluation purposes, only detected contacts are considered in the performance comparison.

Results for the experiments explained in Section 2.4.2 are given separately for the Weston sound speed profiles $P$, the sediment types $S$, and the depths $D$.

## 3.1 Weston sound speed profiles

An example ROC curve for Weston 1 is provided in Figure 2. This figure shows separate ROC curves for the SVM, FCN, and CNN machine learning systems. For all three machine learning systems, the zero-shot / generalization approach has the worst performance, although for the FCN and CNN it is considerably better than for the SVM. The few-shot and environment-specific approaches show the best performance for all three machine learning approaches. For the FCN and CNN, the few-shot approach performs best.

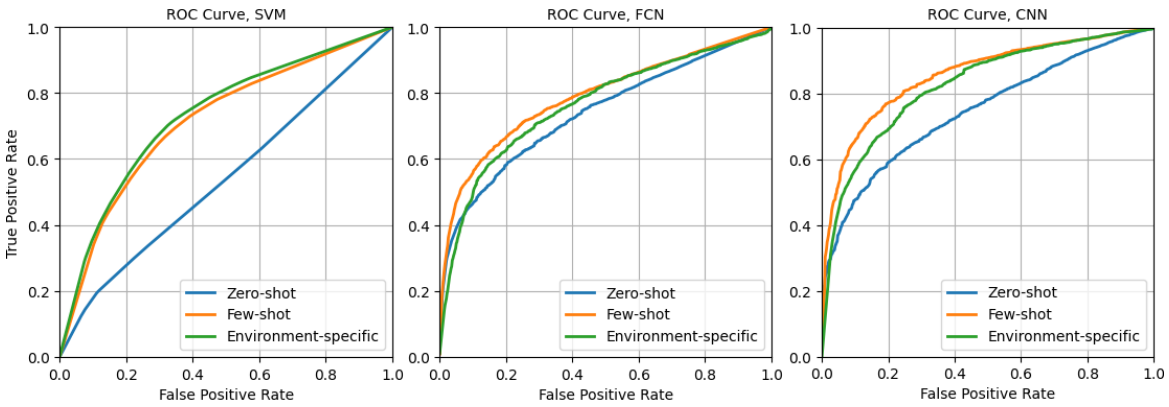


*Figure 2 Receiver operator characteristics (ROC) curves for the Weston 1 sound speed profile, for the SVM (left), FCN (middle), and CNN (right). Overall, the zero-shot / generalization approach performs worse than the few-shot and environment-specific approaches. The CNN appears to outperform SVM and FCN, which is also confirmed by the higher AUC values in Table 1.*

The results for all Weston sound speed profiles are summarized in Table 1, which presents the AUC values. It can be observed that the classification results in Weston 7 are better than in the other Weston profiles, likely because it is easier to classify objects in deeper water (this trend is also observed in the results where depth is varied). The reason behind this could be that the multipath is more spread out in time in these environments. The first and second arrivals happen at distinctly different times and have no overlap, compared to the shallow environments where the differences in path lengths are very small.

*Table 1 Receiver operator characteristics (ROC) area-under-the-curve (AUC) values for all experiments in which the Weston sound speed profile was varied.*

| | **Zero-shot** | | | **Few-shot** | | | **Environment-specific** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test set** | SVM | FCN | CNN | SVM | FCN | CNN | SVM | FCN | CNN |
| Weston 1 | 0.54 | 0.74 | 0.75 | 0.72 | 0.79 | 0.86 | 0.73 | 0.77 | 0.83 |
| Weston 2 | 0.56 | 0.70 | 0.69 | 0.69 | 0.76 | 0.85 | 0.70 | 0.79 | 0.85 |
| Weston 3 | 0.56 | 0.71 | 0.74 | 0.68 | 0.75 | 0.85 | 0.67 | 0.73 | 0.80 |
| Weston 7 | 0.60 | 0.80 | 0.83 | 0.84 | 0.91 | 0.93 | 0.83 | 0.88 | 0.90 |

### 3.2 Sediment types

The results for the various sediment types are summarized in Table 2. It can be observed that a more clay-like sediment type in the environment makes classification less complicated. This is because the rays loose a significant amounts of energy when they interact with the sediment, meaning that there is very little multipath as only the direct arrivals remain. In general, the more sandy the sediment type is, the harder classification is.

*Table 2 Receiver operator characteristics (ROC) area-under-the-curve (AUC) values for all experiments in which the sediment type was varied.*

| | Zero-shot | | | Few-shot | | | Environment-specific | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test set** | SVM | FCN | CNN | SVM | FCN | CNN | SVM | FCN | CNN |
| Coarse clay | 0.63 | 0.83 | 0.82 | 0.82 | 0.90 | 0.92 | 0.82 | 0.89 | 0.92 |
| Medium silt | 0.58 | 0.75 | 0.75 | 0.73 | 0.85 | 0.88 | 0.74 | 0.80 | 0.89 |
| Medium sand | 0.58 | 0.74 | 0.71 | 0.69 | 0.79 | 0.81 | 0.69 | 0.75 | 0.81 |
| Very coarse sand | 0.59 | 0.73 | 0.72 | 0.69 | 0.76 | 0.80 | 0.69 | 0.72 | 0.81 |

### 3.3 Depths

The results for the various depths are summarized in Table 3. It can be observed that the most shallow environment and the deep environments are least complicated environments when it comes to classification. For the deep environments, this is likely for the reasons mentioned before, since a depth larger than 500 m was only simulated in combination with Weston 7. For the most shallow environment, the reason behind this could be that multipath rays have many interaction with the surface and bottom, causing these rays to loose a significant amount of energy. This means that there is less multipath since only the direct arrivals remain.

*Table 3 Receiver operator characteristics (ROC) area-under-the-curve (AUC) values for all experiments in which the depth was varied.*

| | Zero-shot | | | Few-shot | | | Environment-specific | | |
|---|---|---|---|---|---|---|---|---|---|
| **Test set** | SVM | FCN | CNN | SVM | FCN | CNN | SVM | FCN | CNN |
| Depth 50 m | 0.57 | 0.72 | 0.74 | 0.77 | 0.73 | 0.90 | 0.78 | 0.88 | 0.92 |
| Depth 100 m | 0.57 | 0.69 | 0.68 | 0.67 | 0.72 | 0.78 | 0.69 | 0.74 | 0.81 |
| Depth 150 m | 0.58 | 0.65 | 0.68 | 0.65 | 0.67 | 0.79 | 0.67 | 0.71 | 0.79 |
| Depth 200 m | 0.59 | 0.79 | 0.80 | 0.69 | 0.82 | 0.89 | 0.68 | 0.72 | 0.83 |
| Depth 250 m | 0.59 | 0.75 | 0.80 | 0.70 | 0.81 | 0.88 | 0.69 | 0.73 | 0.79 |
| Depth 500 m | 0.63 | 0.89 | 0.88 | 0.82 | 0.91 | 0.94 | 0.80 | 0.89 | 0.92 |
| Depth 1000 m | 0.70 | 0.79 | 0.86 | 0.84 | 0.91 | 0.97 | 0.81 | 0.91 | 0.94 |
| Depth 1500 m | 0.73 | 0.84 | 0.94 | 0.85 | 0.88 | 0.90 | 0.82 | 0.85 | 0.89 |
| Depth 2000 m | 0.70 | 0.95 | 0.92 | 0.86 | 0.98 | 0.97 | 0.84 | 0.92 | 0.92 |
| Depth 2500 m | 0.69 | 0.91 | 0.93 | 0.84 | 0.91 | 0.95 | 0.85 | 0.88 | 0.94 |

## 4 DISCUSSION AND CONCLUSION

This work has focused on the optimal simulation and training strategy of machine learning systems to classify underwater contacts. Given the variability of echoes with position (range, direction, depth) and propagation conditions, this results in environment-dependent differences in the data that commonly used zero-shot approaches cannot deal with. Experiments with simulated data were conducted, confirming that the use of in-situ data in a few-shot or environment-specific setting improves the performance of various machine learning systems.

Some limitations should be considered regarding this work. The simulated targets and environments are theoretical and lack imperfections observed in measured data. In addition, the calculated background levels lack realistic range-dependency characteristics. Furthermore, the machine learning methods used in this study are relatively simple and more sophisticated approaches could possibly achieve better results. Nevertheless, these limitations likely do not affect the overall conclusion that inserting in-situ data in the training dataset is beneficial.

The CNN machine learning systems consistently outperforms the SVM and FCN approaches. The performance difference between the FCN and CNN is relatively small and possibly caused by the number of trainable parameters; the FCN being a much larger network that is harder to optimize with e.g. the limited data in the environment-specific settings. Other neural network architectures might be considered (e.g. transformers) that might make better use of the global echo features present in the snippets when compared a (more local approach of a) CNN.

In conclusion, the zero-shot / generalization approach showed the worst performance, as expected. This suggests that systems do not generalize to unseen environments and including (some) in-situ data of new environments is needed for good performance. In practice, in-situ simulated training data can be obtained by using a target echo simulation and injection capability, enabled by the availability of environment information. The few-shot approach shows slightly better performance than the environment-specific approach, with the trade-off that it requires more data. Both approaches should be considered when designing machine learning systems for underwater target classification.

# 5 ACKNOWLEDGMENTS

This work was funded by TNO's Appl.AI program and the Netherlands Ministry of Defence.

# 6 REFERENCES

[1] D. Abraham, Underwater Acoustic Signal Processing, Ellicott City, MD, USA: Springer Nature Switzerland, 2019.

[2] S. P. Beerens and W. Boek, "A robust algorithm for LFAS target classification," in *Proceedings of UDT Europe*, 2007.

[3] G. De Magistris, P. Stinco, J. R. Bates, J. M. Topple, G. Canepa, G. Ferri, A. Tesei and K. Le Page, "Automatic object classification for low-frequency active sonar using convolutional neural networks," in *OCEANS 2019 MTS/IEEE SEATTLE*, 2019.

[4] P. Stinco, A. Tesei and K. D. LePage, "Unsupervised active sonar contact classification through anomaly detection," in *EURASIP Journal on Advances in Signal Processing*, 2023.

[5] K. T. Hjelmervik, H. Berg, D. H. S. Stender, W. Oxholm and T. S. Såstad, "Synthesizing active anti-submarine warfare sonar data," in *OCEANS 2019-Marseille*, 2019.

[6] M. Ainslie, "Editorial: Validation of Sonar Performance Assessment Tools, in Validation of Sonar Performance Assessment Tools (workshop held in memory of David E. Weston, 7-9 April 2010)," in *Proceedings of the IOA*, 2010.

[7] M. Zampolli, M. Ainslie and P. Schippers, "Scenarios for Benchmarking Range-Dependent Active Sonar Performance Models," in *Validation of Sonar Performance Assessment Tools, in memory of David E. Weston, 7–9 April 2010, Cambridge, UK*, 2010.

[8] M. Ainslie, Principles of sonar performance modelling, Berlin: Springer, 2010.

[9] I. Hartstra, M. Colin and M. Prior, Active sonar performance modelling for Doppler-sensitive pulses, vol. 44, Acoustical Society of America, 2021, p. 022001.

[10] W. Fender, "Scattering from an elastic spherical shell," Naval Undersea Center, San Diego, California, 1972.

[11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg and others, “Scikit-learn: Machine learning in Python.,” in *The Journal of machine Learning research*, 2011.

[12] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga and others, “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in neural information processing systems*, 2019.

[13] W. A. Falcon, “Pytorch lightning,” *GitHub,* vol. 3, 2019.