

RESEARCH

Open Access



# Unsupervised active sonar contact classification through anomaly detection

Pietro Stinco<sup>1\*</sup> , Alessandra Tesei<sup>1</sup> and Kevin D. LePage<sup>1</sup>

\*Correspondence:  
[pietro.stinco@cmre.nato.int](mailto:pietro.stinco@cmre.nato.int)

<sup>1</sup> NATO STO - Centre for Maritime  
Research and Experimentation  
(CMRE), Viale San Bartolomeo  
400, La Spezia, Italy

## Abstract

Target detection and sonar contact classification with active sonar systems are not trivial especially when operating in coastal and shallow water environments with multipath propagation, high reverberation and clutter. It is even more difficult when the sonar receiver is hosted on unmanned platforms with limited maneuvering capabilities unable to perform long-lasting tracking procedures. In such environments with high clutter density, real-time classification algorithms to discriminate target contacts from clutter contacts become crucial. This paper describes a method for active sonar clutter classification that exploits the large number of undesired contacts to learn the “fingerprint” of the environmental clutter and thus to identify the target contacts as anomalies. The paper introduces the method to obtain the features of detected sonar contacts from the beamformed signal of a triplet array of hydrophones that can be towed by an autonomous underwater vehicle. The paper also shows the performance of the proposed unsupervised classification algorithm with real data collected at sea and compares it to what has been achieved by using a convolutional neural network. The results show the capability of the proposed anomaly detection algorithm to properly deal with a variety of clutter contacts without requiring labeling of training data.

**Keywords:** SONAR, Automatic target classification, Unsupervised learning, Anomaly detection

## 1 Introduction

Long range detection of underwater targets is increasingly performed using low-frequency active sonar (LFAS) systems which consist of a powerful wideband source and one or more receiving hydrophone arrays.

The receiving arrays are typically towed by vessels [1–3] or, especially when operating in coastal environments, by autonomous underwater vehicles (AUVs). AUVs are characterized by lower sensing, computational and communication capabilities as compared to traditional assets. However, they can build intelligent networks to accomplish complex missions with features of redundancy, persistence, scalability and adaptability.

This opens the possibility to develop a multistatic network of cooperative and autonomous platforms for underwater surveillance, for operations in littoral, shallow waters environments that have become more and more important in many civilian and defense applications [4, 5].

In such environments, multipath propagation, high reverberation and clutter are the main challenges faced by LFAS systems.

An important solution to mitigate the effect of reverberation and clutter on the detection performance is the use of an array of advanced directional sensors, such as twin arrays [6–8], acoustic probes [9–14], acoustic vector sensors [15–18] or triplet arrays [19–23], an example of which is exploited in this work.

These types of sensors are able to solve bearing ambiguity in a single ping and to reject coastal reverberation in offshore bearings. This is not possible with single-line array receivers since they are cylindrically symmetric and therefore cannot discriminate port from starboard.

However, when operating in high-density clutter environment, even with directional sensors, the use of classification algorithms to distinguish target contacts (if present) from the large number of clutter contacts to reduce the false alarm rate at an acceptable probability of detection becomes crucial.

Traditionally, sonar contact classification is left to sonar operators who use their expertise to discriminate targets from clutter. Automatic target classification (ATC) can be helpful to reduce the workload on sonar operators and is necessary to reduce false alarms when the receivers are robots operating without human supervision, such as AUVs.

Recently, deep learning classifiers based on convolutional neural networks (CNNs) have been used for underwater target classification with high-frequency sonar images for mine-hunting obtaining very good performances [24].

Mine-hunting sonars work at short ranges of few tens of meters and, hence, can operate at much higher frequencies compared to LFAS. Moreover, the pulse repetition interval is much lower than the coherence time of the insonified object and hence multiple pulses can be coherently processed to extract information on the target. The amount of information bandwidth from these wideband scans is sufficiently large not only for mine-clutter discrimination but also for mine type recognition.

In LFAS systems, the coherence time of target is much lower than the pulse repetition interval and hence only one pulse can be processed for target detection and classification. Moreover, the bandwidth LFAS pulses is much lower than the one exploited by mine-hunting sonar. Even with these limitations, in recent results [25, 26] we demonstrated that CNNs provide good classification performance for ATC. However, there are a number of challenges to face when using supervised learning techniques in LFAS systems.

First of all, extensive data collection with targets is extremely costly, it is difficult to generate accurate datasets for training, and data labeling is not straightforward and usually very time-consuming.

Moreover, the collected training datasets are often unbalanced and consist of large amounts of clutter data with very few target data for each aspect angle. With such dataset and constraints, it is difficult to train a general model for ATC.

It is also worth to consider that the navigation accuracy of underwater vehicles and the azimuthal localization accuracy of towed arrays are generally not high, and the estimated position of a target contact is often far away from the real one. Moreover, the

target is often sailing very close to cluttered regions (especially in shallow water environments), which significantly increases the probability to erroneously label a clutter contact as a target and vice versa.

Starting from the results introduced in [27], this paper introduces an unsupervised learning method based on an anomaly detection for target classification.

Anomaly detection has been successfully used in many applications, such as in cyber security [28], financial fraud [29], target detection in hyperspectral images [30], mammographic image analysis [31] or mine detection and classification [32, 33], just to name only a few.

There are many approaches to anomaly detection based on statistical models, machine learning, saliency based-methods, sparse representations, and more [34–36].

The method presented here is based on machine learning since the model is learned by training the algorithm with a dataset of clutter samples, the availability of which is much higher than the samples of targets of interest.

The proposed method exploits the clutter contacts to learn the clutter signature and then classifies the target contacts as anomalies if their signature is not similar to the learned one, representative of the natural environment.

In real operations, the receiver can learn the clutter signature during a pre-survey of the operational area. This approach is similar to the one proposed in [37], where the active sonar receiver increases its knowledge of the surrounding environment generating clutter maps based on the persistency of clutter contacts over a geographical area.

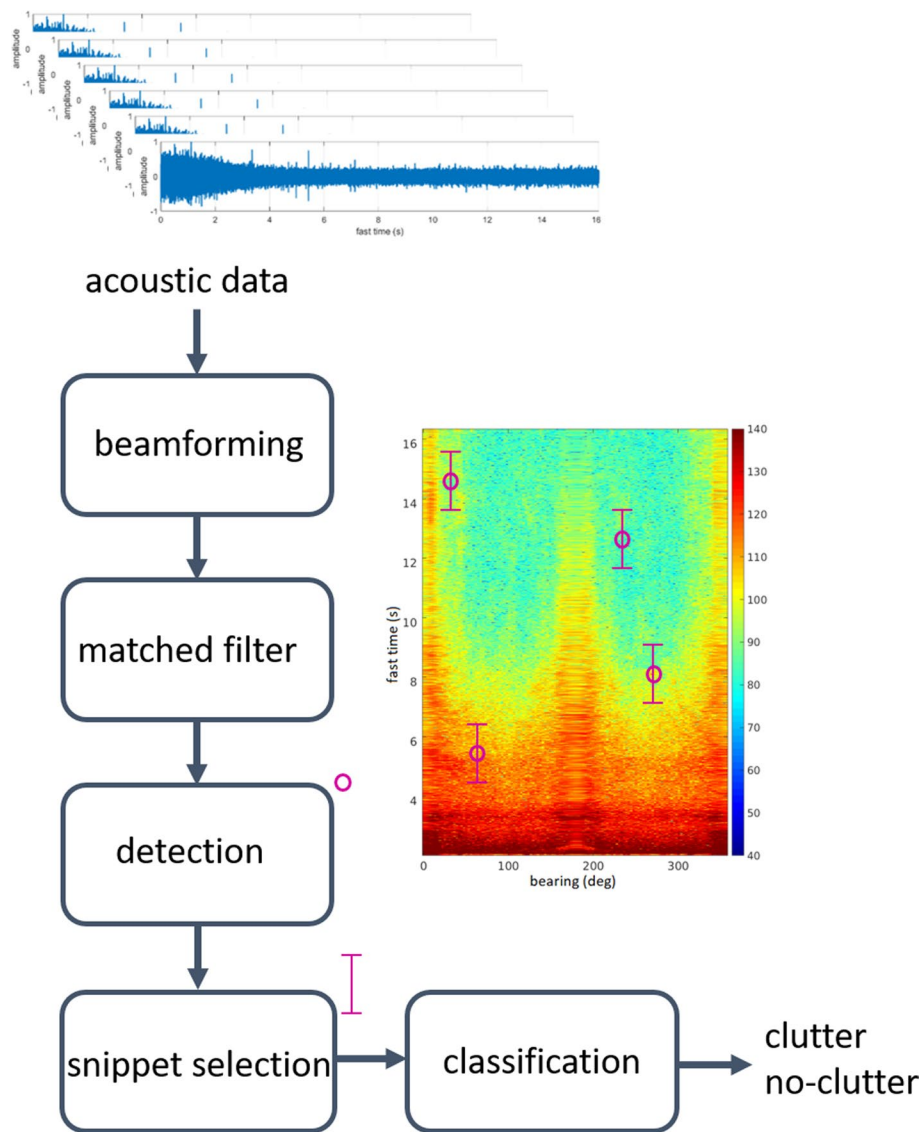
During operations at sea, it is possible to periodically update the clutter signature estimate. This is very important when operating in shallow waters where the underwater environment rapidly changes and it is fundamental for adapting the receiver in order to reduce the false alarm rate.

The first section of this paper introduces the signal processing steps used to obtain the snippets of the sonar contacts from the raw acoustic data at the hydrophones of the array. These snippets are the inputs of the proposed sonar contact classification algorithm described in the second section of the paper. The third section of the paper shows the performance of the proposed unsupervised classification algorithm with real data collected at sea and compares it to what has been achieved by using a CNN trained on a labeled dataset.

The results have been obtained from experimental data collected at sea using an echo-repeater (E/R) as an artificial target and the SLIm Cardioid Towed Array (SLICTA) as the sonar receiver. The SLICTA is a triplet array with port-starboard discrimination capability designed and developed at CMRE to be towed by an Ocean Explorer (OEX) AUV.

## 2 Signal processing before classification

This section summarizes the signal processing steps used to obtain the time snippets on the sonar contacts that are the inputs of the proposed classification algorithm. All these steps are illustrated in Fig. 1.

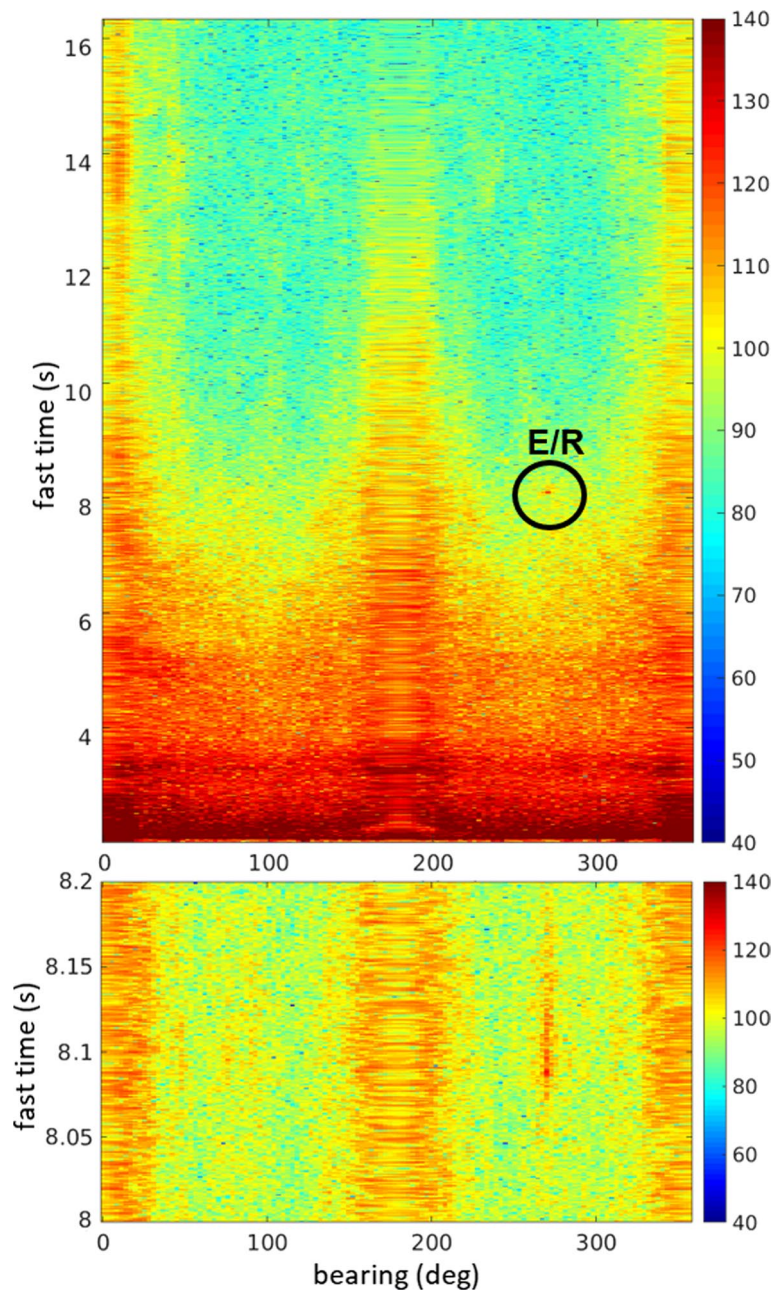


**Fig. 1** Flowchart of the signal processing chain to provide the inputs of the classification algorithm. As an illustrative example, the figure shows the fast time-bearing map at the output of the beamformer and matched filter, the detections (magenta circles) and the time snippets (magenta segments) used for sonar contact classification

In LFAS systems, the acoustic source transmits a frequency modulated sonar pulse with a given pulse repetition interval. For each of these pings, the acoustic signals collected by the hydrophones of the receiving array are beamformed in all the steering directions to obtain the fast time-bearing map of the surveillance area. The transmitted signals are frequency-modulated pulses. The wider the bandwidth, the higher the range resolution and, from physical acoustic, the more the classification features the signal contains. This is in agreement with the Shannon theorem that states that the wider is the bandwidth and the signal-to-noise ratio the more is the amount of information in the received signal.

Typical transmitted signals in LFAS system are linearly frequency modulated pulses (LFM) and hyperbolic frequency modulated (HFM) pulses; however, target classification is performed after matched filtering, and hence, the classification algorithm is independent on the type of frequency modulation.

As an illustrative example, Fig. 2 shows the fast time-bearing map of a single ping. This output shows for each bearing direction the absolute value in dBre $\mu$ Pa of the



**Fig. 2** Top: fast time-bearing map at the output of the adaptive beamformer and matched filter. Signal collected with the E/R at bearing 270° and range 6 km (fast time 8 s). Bottom: zoom on the target echo



received signal between two consecutive transmissions after beamforming and matched filtering.

The vertical axis is the fast time, which is directly proportional to range by a factor of half the speed of sound. The horizontal axis is the bearing direction, rotating positive clockwise from the towing direction at forward end-fire ( $0^\circ$ ).

In this particular example, the data have been collected using the SLICTA triplet array with port-starboard discrimination capability. The bearing angle rotates from  $0^\circ$  to  $360^\circ$ .

The data have been collected using an E/R playing the role of an artificial target. It was at broadside port (bearing  $270^\circ$ ) at a range of about 6000 m, corresponding to a fast time of about 8 s.

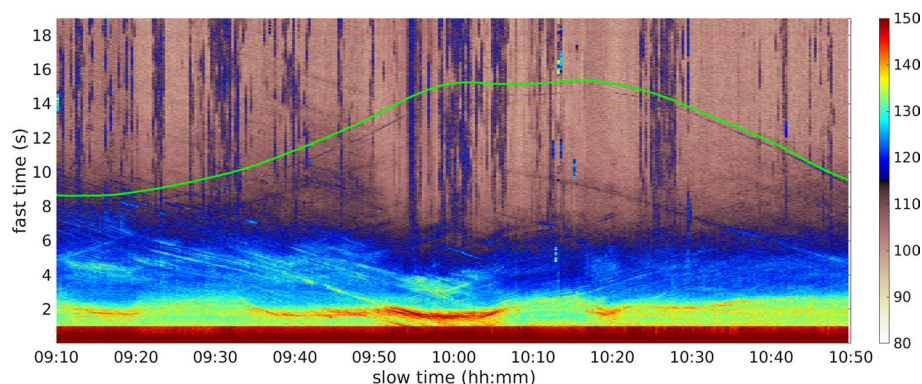
The fast time-bearing map shown in Fig. 2 has been obtained using the adaptive beamformer described in [23], but other beamformers can also be used. The classification algorithm described in this section is valid for any beamformer and any array of hydrophones even without port-starboard discrimination capabilities.

The adaptive beamformer is used since it allows to unmask objects in the presence of strong coastal reverberation and/or traffic noise on other bearings. It is a type of minimum variance distortionless response (MVDR) beam space adaptive algorithm where the inner triplet correlations are actually measured and, for each steering direction and at each range cell, the beamforming is adapted to the local environment. In this way, port-starboard discrimination is guaranteed in beams with directional coastal reverberation, while high signal-to-noise ratios are obtained in offshore (noise limited) beams.

Figure 2 also shows a zoom on the E/R echo, and it is quite evident how the beamformer is able to completely reject the ambiguous echo at starboard (bearing  $90^\circ$ ).

Figure 2 shows the output of a single ping of a run with duration of about 2 h and whose beam collapse plot (BCP) is shown in Fig. 3.

This dataset has been collected in the monostatic configuration with both the receive SLICTA array and the acoustic source towed by the NATO Research Vessel (NRV) Alliance. The Coastal Research Vessel (CRV) Leonardo was towing the E/R to play the role of the artificial target.



**Fig. 3** Beam collapse plot at the output of the adaptive beamformer and target ground truth (green) delayed of 0.5 s

As suggested by its name, the BCP shows the output of the beamformer for all the pings of the run [23]. It shows, for each ping (slow time) and each fast time point, the maximum value of the beamformer output along the bearing direction. The green line in Fig. 3 is the ground truth that is the expected round trip delay between target and receiver. This line has been delayed of 0.5 s to avoid overlap with the echoes from the E/R, which are evident in the dark purple color.

It is evident how the E/R signal is embedded in a highly cluttered environment.

Very strong echoes are coming from the seafloor. Compact clutter is quite visible even at very far ranges. In this run, the receiver is approaching a quite large region of compact clutter which is at about 6–10 s at the beginning of the run and very close to the receiver at about 10:00. Reverberation is also quite strong in the first 8 s of the ping, and the signal is also affected by interfering ship traffic noise that mainly consists of the continuous signals clearly visible in the BCP as vertical stripes.

All these non-target signals provide information on the operational environment and can be exploited to estimate the clutter signature that is the “fingerprint” of the area where the active sonar is operating.

After beamforming and matched filtering, detection is performed to find the fast time-bearing coordinates of the contacts that must be processed for target-clutter discrimination.

In the receiver described in this paper, detection is performed on the fast time-bearing map with the Ordered Statistics-Constant False Alarm Rate (OS-CFAR) detector [38], which is able to detect objects embedded in background with an unknown and non-stationary statistical distribution.

Also in this case, the classification algorithm is independent on the detector used to find the contacts to be classified.

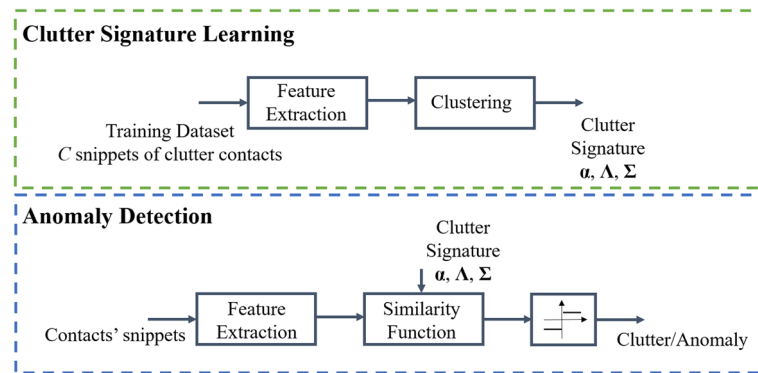
The coordinates of each detected object are then used to extract a snippet from the fast time-bearing map at the output of the beamformer. The snippet is the matched filtered time series centered at the contact’s range-bearing coordinates; this is the input of the classification algorithm described in next section.

As discussed in the Introduction, beamforming before classification is fundamental to separate the contacts in the space domain. It is also possible to perform classification on the signal collected by a single hydrophone but, in this case, contacts at the same distance but at different bearing directions are superimposed, and hence, it is impossible to separate the features of each single contact, with a resulting degradation of the classification performance.

### 3 Sonar contact classification

The block diagram of the unsupervised active sonar contact classification method proposed in this paper is shown in Fig. 4.

The figure shows the two phases of the anomaly detection approach for object classification. In the first phase, the algorithm learns the clutter signature by processing the features extracted by the acoustic signal of a training set containing only clutter contacts.



**Fig. 4** Block diagram of the classification algorithm. *Clutter signature learning*: learn the clutter signature from a training set of clutter contacts. *Anomaly Detection*: contact classification by comparing the signature of each contact with the learned clutter signature in target-free environment

In the second phase, the features of each sonar contact are compared with the learned clutter signature and then an object is detected if its features are anomalous, i.e., not similar to the learned one.

### 3.1 Contact features

We indicate with  $s_c[n]$  the snippet on the contact  $c$ . The snippet is the finite time sequence with the samples of the contact at a certain beam, i.e., the output of the beam-former at the time-bearing coordinates of the detection.

Let us also indicate with  $\mathbf{f}(s_c[n])$ , in short  $\mathbf{f}_c$ , a vector that collects  $M$  real valued features on the contact  $c$ , i.e.,

$$\mathbf{f}_c = \mathbf{f}(s_c[n]) \in \mathbb{R}^M \quad (1)$$

Generally speaking, the size  $M$  of the feature vector is of the order of few tens (30–40 or even more) and the numerical features are scaled in order to have values of the same order of magnitude (for example, with values from  $-1$  to  $1$ ). A good way to choose the contact features is to exploit those that might take unusually large or small values in the event of an anomaly.

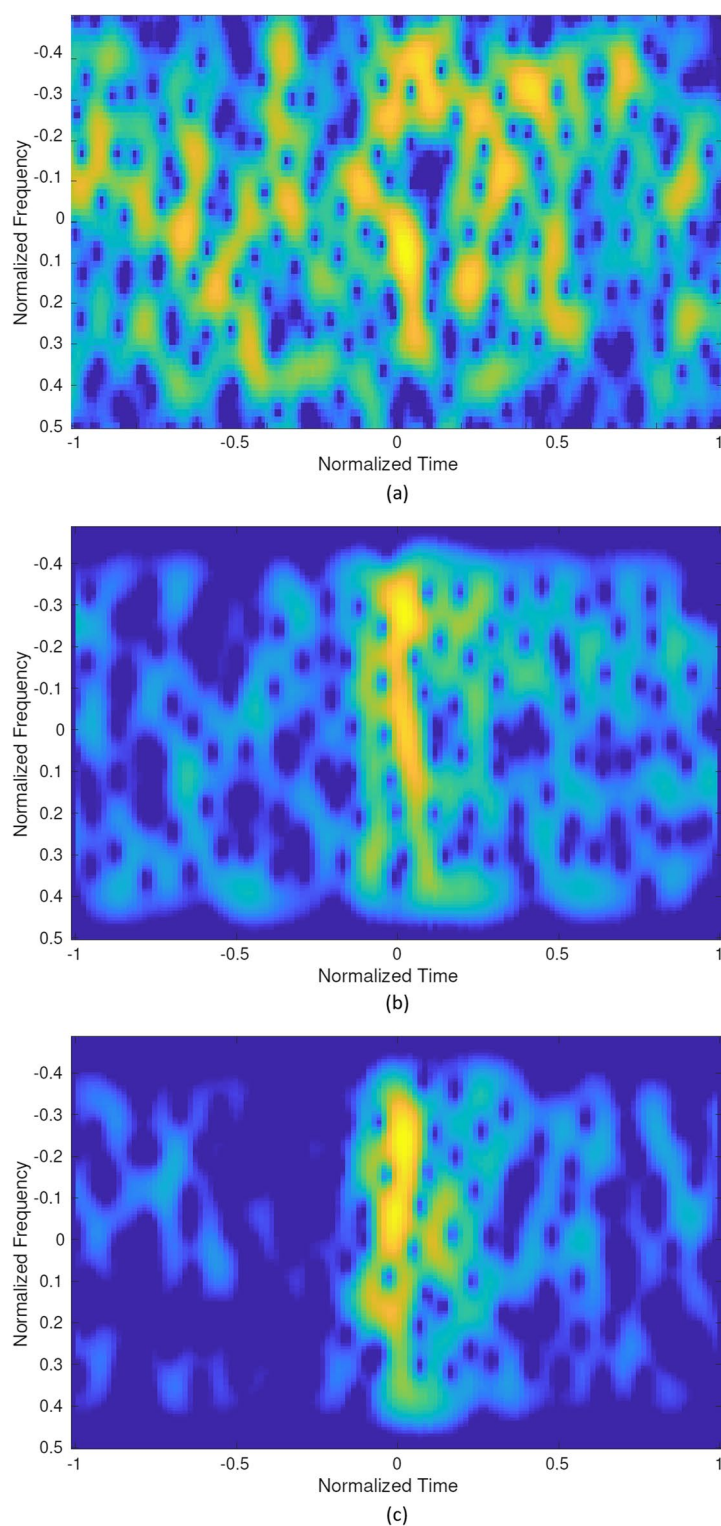
In this work, all the features are extracted starting from the spectrogram  $S_c(t, f)$  of the acoustic time snippet  $s_c[n]$ .

The spectrogram shows how the power of the acoustic echoes is distributed over time at various frequencies and hence can be used for sonar contacts discrimination.

The spectrogram is the bi-dimensional image obtained by dividing the time snippet into subsequences with maximum overlap, multiplying each subsequence with a Hamming window and by taking the squared absolute value of the fast Fourier transform (FFT) of each windowed subsequence.

Figure 5 shows the spectrograms of three different sonar contacts. In the best case scenario, the clutter response is typically uniformly spread in the analyzed time–frequency





**Fig. 5** Spectrograms of **a** clutter contact, **b** ambiguous contact and **c** target contact (echo-repeater)

window (see Fig. 5a), while the target (echo-repeater) is more focused in the central part of the spectrogram (normalized time = 0) and/or in the right portion of the spectrogram (normalized time > 0), as in Fig. 5c. Figure 5b shows the spectrogram of a typical clutter contact that can be easily misclassified and then labeled as a target.

All the elements of the feature vector are extracted from the spectrogram. In our specific case, the feature vector size  $M$  is 32.

The first five elements of  $\mathbf{f}_c$  are the first five raw time moments of the spectrogram, i.e.,

$$m_t(k) = \frac{1}{F} \frac{1}{T^k} \int_{-F/2}^{F/2} \int_{-T/2}^{T/2} S_c(t, f) t^k dt df \quad (2)$$

with  $k = 1, 2, \dots, 5$ .

Similarly, the subsequent five elements are the first five raw frequency moments, i.e.,

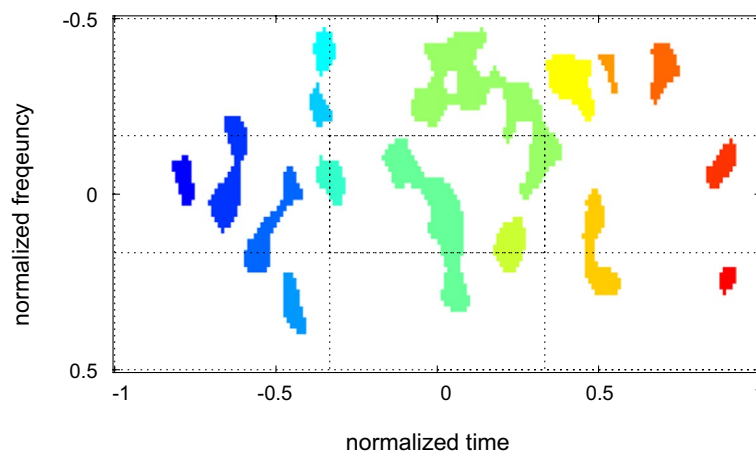
$$m_f(k) = \frac{1}{F^k} \frac{1}{T} \int_{-F/2}^{F/2} \int_{-T/2}^{T/2} S_c(t, f) f^k dt df \quad (3)$$

with  $k = 1, 2, \dots, 5$ .

It is possible to exploit also higher-order moments; however, for a given number of pixels in the spectrogram, the higher the moment's order, the higher the variance of its estimate. In our case,  $k = 5$  is a good compromise, since commonly adopted values of  $k$  are 3–4, that is, the same order of skewness and kurtosis, respectively.

The remaining elements of the feature vectors are derived from statistics computed on the blobs.

The blobs are the output of a conventional flood-fill connected-component clustering algorithm [39] applied to the spectrogram. As an illustrative example, Fig. 6 shows the blobs of the spectrogram in Fig. 5a. In this particular case, there are 16 blobs, each identified by a different color.



**Fig. 6** Blobs of the spectrogram in Fig. 5a

---

**Algorithm 1** Clutter Signature Learning
 

---

**Input:** features of clutter contacts  $\mathbf{F} = [\mathbf{f}_1 \dots \mathbf{f}_C]$   
**Output:** clutter signature  $\alpha, \mathbf{\Lambda}, \mathbf{\Sigma}$

- 1: randomly initialize  $N$  centroids  $\theta_i$  from  $\mathbf{F}$
- 2: randomly initialize  $c$  ( $c_i$  - index to which  $\mathbf{f}_i$  is assigned)
- 3: initialize  $d = 0$
- Clustering*
- 4: **while** ( $d \neq c$ ) **do**
- 5:    $d = c$
- 6:   **for** ( $i = 1 : C$ ) **do**
- 7:      $c_i \leftarrow$  closest to  $\mathbf{f}_i$  centroid using  $\text{dist}(\mathbf{f}_i, \theta_i)$
- 8:   **end for**
- 9:   **for** ( $i = 1 : N$ ) **do**
- 10:      $\theta_i \leftarrow$  average of all points assigned to  $c_i$
- 11:   **end for**
- 12: **end while**
- Compression*
- 13:  $k = 0$
- 14:  $N_\alpha = 0$
- 15: **for** ( $i = 1 : N$ ) **do**
- 16:    $n_i \leftarrow$  number of points with label  $i$
- 17:   **if** ( $n_i > n_{min}$ ) **then**
- 18:      $k \leftarrow k + 1$
- 19:      $\mathbf{X} \leftarrow$  set of points from  $\mathbf{F}$  with label  $i$
- 20:      $\mathbf{X} \leftarrow$  select 80% of closest points  $\mathbf{x}$  to  $\theta_i$
- 21:      $\alpha_k \leftarrow$  number of points in  $\mathbf{X}$
- 22:      $N_\alpha \leftarrow N_\alpha + \alpha_k$
- 23:      $\lambda_k \leftarrow$  centroid of  $\mathbf{X}$
- 24:      $\sigma_k \leftarrow$  average value of  $|\mathbf{x} - \lambda_k|$
- 25:   **end if**
- 26: **end for**
- 27:  $\alpha \leftarrow \alpha / N_\alpha$
- 28: **return**  $\alpha, \mathbf{\Lambda}, \mathbf{\Sigma}$

---

The 11<sup>th</sup> feature is the number of blobs divided by 100. The features from 12 to 23 are statistics on the blobs' size, i.e., standard deviation, mean, minimum and maximum values in both time and frequency domains.

The last 9 elements come from the statistics on the blobs' distribution. The time–frequency window of the spectrogram is divided into 9 sectors, as depicted with the dashed-black lines in Fig. 6, and each feature counts how many blobs fall inside (divided by the total number of blobs). The position of each blob is the time–frequency coordinate of its center of mass.

All the elements of the feature vector are summarized in Table 1.

### 3.2 Clutter signature learning

Let us consider a training dataset consisting of  $C$  clutter contacts. This dataset can be obtained in post-processing by randomly selecting  $C$  contacts at a sufficient distance from the ground truth. In real operations, the training contacts can be collected during a pre-survey of the operational area, under the assumption that all the collected contacts are from clutter.

**Table 1** Elements of the feature vector

Feature vector	
1	1st raw time moment of spectrogram
2	2nd raw time moment of spectrogram
3	3th raw time moment of spectrogram
4	4th raw time moment of spectrogram
5	5th raw time moment of spectrogram
6	1st raw frequency moment of spectrogram
7	2nd raw frequency moment of spectrogram
8	3th raw frequency moment of spectrogram
9	4th raw frequency moment of spectrogram
10	5th raw frequency moment of spectrogram
11	Normalized number of blobs
12	Mean time–frequency size of blobs
13	Minimum time–frequency size of blobs
14	Maximum time–frequency size of blobs
15	Standard deviation of blobs' time–frequency size
16	Mean time duration of blobs
17	Minimum time duration of blobs
18	Maximum time duration of blobs
19	Standard deviation of blobs' time duration
20	Mean frequency band of blobs
21	Minimum frequency band of blobs
22	Maximum frequency band of blobs
23	Standard deviation of blobs' frequency band
24	Number of blobs in the 1st sector
25	Number of blobs in the 2nd sector
26	Number of blobs in the 3th sector
27	Number of blobs in the 4th sector
28	Number of blobs in the 5th sector
29	Number of blobs in the 6th sector
30	Number of blobs in the 7th sector
31	Number of blobs in the 8th sector
32	Number of blobs in the 9th sector

The algorithm consists of two phases: a first *clustering* phase for an initial fit of the features model and a following *compression* phase to remove clusters with few samples and to filter out possible outliers that can affect the clutter signature estimate. This is done in real operations at sea to avoid that the presence of unexpected target contacts in the operational area can affect the clutter characterization.

Clustering is performed using the *K-means* method. The algorithm is initialized by randomly selecting  $N$  centroids from the input feature points, where  $N$  is of the same order of magnitude as  $M$ . The clustering algorithm iteratively assigns a label  $c_i$  to any feature point, considering the nearest centroid and using the Euclidean distance.

After that, a label is assigned to every feature point, and the values of the cluster centroids are updated by taking the average over all points with the same label.

The algorithm stops when the labels no longer change.

The second step of the learning algorithm is the compression phase. In this phase, the clusters with few points are discarded. The minimum number of points is a small fraction (1–5%) of  $C/N$ . Moreover, for each of the remaining  $K \leq N$  clusters, all the points far away from their centroid are discarded. For each cluster, only the 80% of points closest to the centroid are processed to estimate the clutter signature. This is done to avoid that, during the training dataset collection in the pre-survey of the operational area, some anomalies, such as the presence of unexpected targets, can affect the clutter signature estimate.

The clutter signature is described by the three quantities

$$\begin{aligned}\alpha &\in \mathbb{R}^K \\ \Lambda &= [\lambda_1 \dots \lambda_k \dots \lambda_K] \in \mathbb{R}^{M \times K} \\ \Sigma &= [\sigma_1 \dots \sigma_k \dots \sigma_K] \in \mathbb{R}^{M \times K}\end{aligned}\quad (4)$$

where  $\alpha$  is a vector collecting the normalized number of points (normalized with the total number of processed points) of each cluster,  $\Lambda$  is a  $M \times K$  matrix that collects in each column the centroids of the clusters (after outliers removal), and  $\Sigma$  is a  $M \times K$  matrix whose  $k$ th column is the element-by-element mean distance of the points from the  $k$ th centroid.

### 3.3 Anomaly detection

Once the clutter signature has been learned, the receiver evaluates, for each contact  $c$ , the features vector  $\mathbf{f}_c$  as in (1) and then computes the similarity function  $\rho_c$  using

$$\rho_c = \sum_{k=1}^K \Theta_k(\mathbf{f}_c) \quad (5)$$

where  $\Theta_k(\mathbf{f}_c)$  is the Gaussian kernel of cluster  $k$

$$\Theta_k(\mathbf{f}_c) = \alpha_k \prod_{m=1}^M e^{-\frac{(\mathbf{f}_c(m) - \lambda_k(m))^2}{2\sigma_k^2(m)}}. \quad (6)$$

For each of the  $K$  centroids, the kernel function evaluates whether the features vector is close to the centroid.

A high value of the kernel indicates a high similarity between the features vector and the cluster. The similarity value  $\rho_c$  is given by the sum of the  $K$  kernels, weighted with  $\alpha_k$ . This means that the high-populated clutter clusters have a high contribution in the similarity function. Note that the value of  $\rho_c$  is positive and less than 1,  $\rho_c \in [0, 1]$ .

Notice also from (6) that the feature components are treated as independent, the kernel is given by the product of the similarity between each component of the feature and the coordinates of the centroid. It is also possible to use multivariate Gaussian kernels to evaluate the similarity function by inferring the correlation between the clutter features.

The use of multivariate Gaussian kernels automatically captures and exploits correlations between features, but is computationally more expensive and does not scale with large number of features. This is because, for large number of features  $M$ , it is possible that some of them are linearly dependent. In this case, the resulting covariance matrix is rank deficient and its inversion generates numerical issues.



On the other hand, the use of independent kernels is computationally cheaper, scales better with large features and provides the freedom to design very large vectors of features without taking care of possible dependencies among them.

The last step of the algorithm is contact classification: A contact is labeled as target ( $l_c = 1$ ) if the similarity function is lower than a threshold  $\epsilon$ ; otherwise, it is labeled as clutter ( $l_c = 0$ ), i.e.,

$$l_c = \begin{cases} 1 & \text{if } \rho_c < \epsilon \\ 0 & \text{if } \rho_c \geq \epsilon \end{cases} \quad (7)$$

$\epsilon$  is a very low value that decreases with increasing number of features, and typical values are close to  $10^{-M}$ .

#### 4 Results and performance

Figure 7 shows with dots all the sonar contacts at the output of the sonar receiver for the run described in Sect. 1. The plot shows the measured fast time of all the detections as a function of slow time, compared with the ground truth (solid magenta line). The resulting output consists of 2098 contacts with a rate of almost seven contacts per ping.

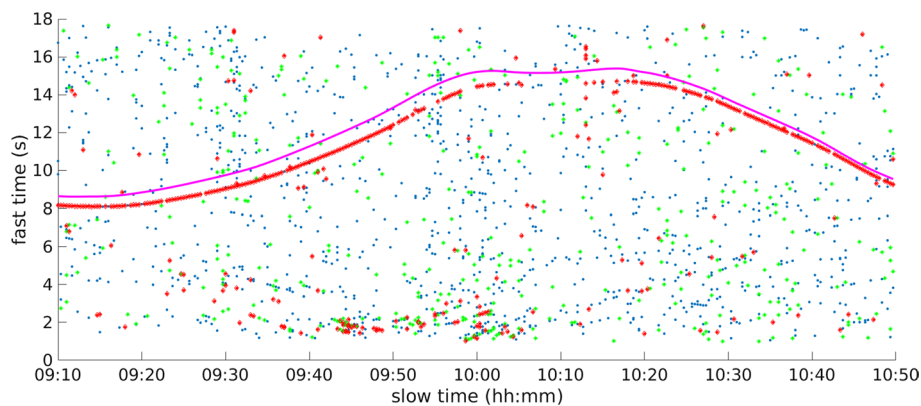
The figure also shows the output of the proposed classification algorithm.

The green dots are the clutter contacts used in the training phase to learn the clutter signature, while the red dots are the anomalous contacts classified as targets. The remaining blue dots are those classified as clutter.

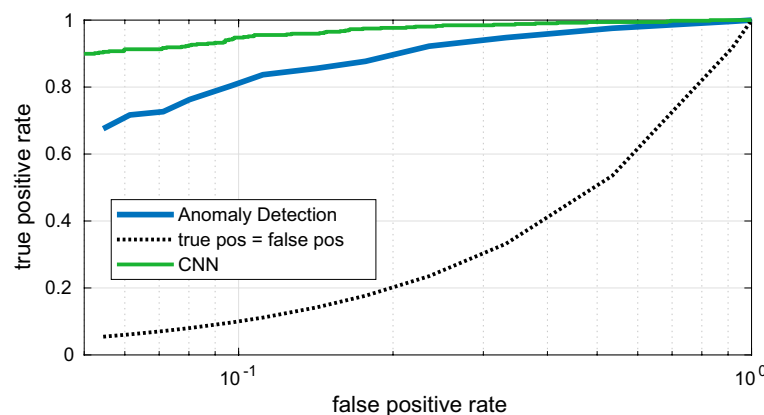
In this example, the green dots are 800 randomly selected clutter contacts representing different kinds of sonar clutter, since, as apparent from the BCP in Fig. 3, they are generated from compact clutter, diffuse reverberation and interfering ship noise.

The anomalies have been obtained with threshold  $\epsilon = 10^{-40}$ ; the resulting percentage of true positives (target contacts correctly classified as target) is close to 80%, while the percentage of false positives (clutter contacts classified as target) is close to 10%.

The performance in terms of receiver operating characteristic (ROC) for the proposed classification method is shown by the blue curve in Fig. 8.



**Fig. 7** Contacts at the output of the signal processing chain compared with the ground truth (magenta). Blue: all sonar contacts. Green: randomly selected clutter contacts exploited for learning clutter signature. Red: anomalous contacts classified as target



**Fig. 8** ROC of the anomaly detection algorithm (blue) compared with the ROC of the CNN in [25] (green); black: worst case where the true positive rate equals the false positive rate

The ROC is the true positive rate as a function of the false positive rate. This plot has been obtained by Monte Carlo runs, averaging the performance obtained with different training sets of 800 contacts and different values of the threshold  $\epsilon$ .

The ROC of the proposed method is compared with the one of the CNN described in [25] (green line in Fig. 8).

As expected, the ROC of the proposed algorithm is lower than the one of the supervised method.

This is mainly due to the fact that the CNN is a supervised classification method that exploits also the information learned from the target.

The ROC of the CNN is excellent but, as discussed in the Introduction and in [25], CNNs need to be trained with large and very accurate datasets, with consequent difficulty and cost in collecting and labeling the data.

In particular, the CNN exploited to get the results in Fig. 8 has been trained with weeks of data collected in two sea trials using an E/R as the target [25].

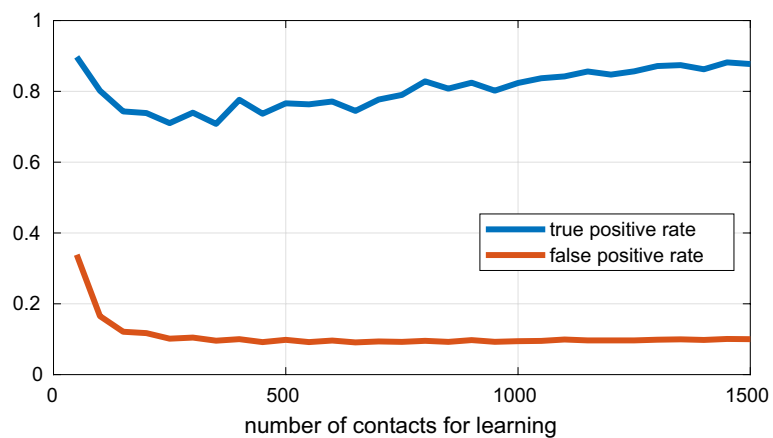
Even if the performance of the unsupervised anomaly detection algorithm is slightly lower than that of the CNN, the algorithm has been trained online by using only few hundreds of clutter samples that can be collected in less than one hour.

For the dataset analyzed, the contacts rate was of almost 20 contacts per minute. Thus, in order to collect a training set of 800 clutter samples, the required time is only 40 min.

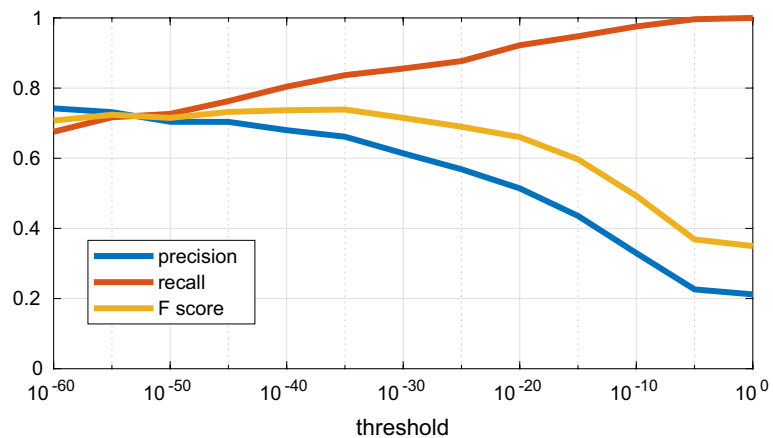
Figure 9 shows the true positive and false positive rates as a function of the number of clutter samples used for training.

From this plot, it is evident how, for this dataset, it is possible to get good performance with few hundred training clutter samples. The higher the number of training samples the higher the true positive rate. Figures 10 and 11 show the performance in terms of precision, recall and F score as a function of the threshold  $\epsilon$  and the training set cardinality, respectively.

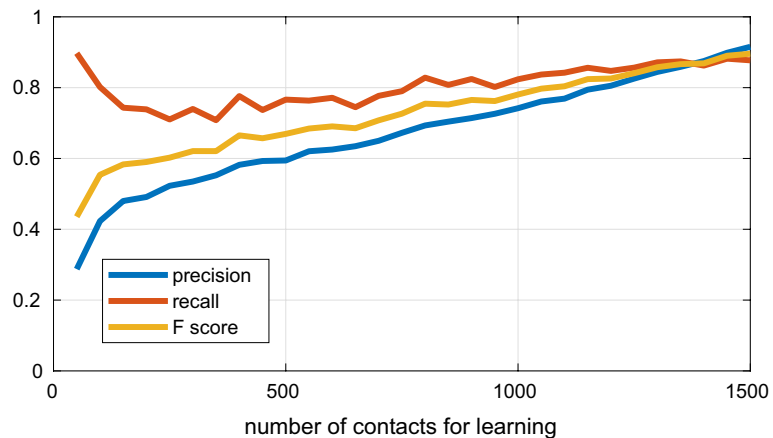
Precision indicates how many positive contacts are relevant and is defined as the rate of true positive among the contacts classified as positive (sum of true positives and false positives). Recall indicates how many relevant contacts are selected and hence is the same as the true positive rate. Recall is defined as the rate of true positives to the sum



**Fig. 9** True and false positive rates as a function of the training set cardinality



**Fig. 10** Precision, recall and F score as a function of the threshold  $\epsilon$



**Fig. 11** Precision, recall and F score as a function of the training set cardinality

of true positives and false negatives. The F score is the harmonic mean of precision and recall.

For increasing values of the threshold, the recall is increasing while the precision is decreasing. This is obvious considering that when the threshold tends to one, all the contacts are classified as anomalies and hence the recall tends to one but the number of false alarms is maximum.

From the plots, precision and recall are similar when the threshold is close to  $10^{-40}$ , with 40 being very close to the number of features of our classifier ( $M = 32$ ). As for the ROC, the higher the number of contacts used to learn the clutter signature, the better the classification performance. Clearly, the larger the training set cardinality, the longer the time required to collect the training set in real operations.

## 5 Conclusions

This paper presented an unsupervised classification method for ATC based on an anomaly detection approach.

The core idea is to take advantage of the huge amount of non-target contacts, especially when operating in littoral waters, to estimate the clutter signature of the operational environment.

This signature can be viewed as the fingerprint of the clutter, and hence, a target can be detected if its features are anomalous: that is not similar to the learned clutter features.

The method is based on machine learning since the clutter model is not predefined but it is learned by training the algorithm with a dataset consisting largely of clutter data.

The main advantage of this algorithm with respect to conventional supervised learning techniques is that there is no need to train the algorithm with target contacts that, especially for underwater surveillance applications, are very difficult and costly to be collected. On the other hand, when operating in challenging littoral, shallow water environment, there is a huge amount of clutter data and the time required to learn online the clutter signature can be less than one hour and, and hence, it is possible to update/refine the clutter signature estimate several times during mission's duration.

This is very important when operating in shallow water environments, where the clutter signature may rapidly change in space and time. A supervised method can easily fail if not trained on the environment being encountered.

The successful performance of the algorithm was demonstrated with real data collected at sea using an echo-repeater as an artificial target.

The results show the capability of the proposed algorithm to cope with a variety of clutter contacts.

Clearly, supervised algorithms such as those based on CNN exploit also the information on target contacts and, hence, have better performance than unsupervised methods.

For this reason, anomaly detection can also be used in post-processing for automatic labeling of sonar contacts, with the aim of building or enriching training sets for the learning phase of supervised classification methods.

Future research will focus on combining the anomaly detection output with those of other supervised and model-based classification algorithms to improve the overall performance of a LFAS system.

**Acknowledgements**

This work has been supported by the NATO Allied Command Transformation under the Autonomy for Anti-Submarine Warfare research program.

**Author contributions**

The authors equally contributed to the paper.

**Funding**

This work has been supported by the NATO Allied Command Transformation under the Autonomy for Anti-Submarine Warfare research program.

**Availability of data and materials**

Please contact author for data requests.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 6 February 2023 Accepted: 23 April 2023

Published online: 31 May 2023

**References**

1. E.C. Whitman, SOSUS: the secret weapon of undersea surveillance. *Undersea Warf.* **7**, 2 (2005)
2. R.O. Nielsen, *Sonar Signal Processing* (Artech House, Norwood, 1991)
3. D.A. Abraham, *Underwater Acoustic Signal Processing: Modeling, Detection, and Estimation* (Springer, Berlin, 2019)
4. G. Ferri, A. Munafó, A. Tesei, P. Braca, F. Meyer, K. Pelekanakis, R. Petroccia, J. Alves, C. Strode, K.D. LePage, Cooperative robotic networks for underwater surveillance: an overview. *IET Radar Sonar Navig.* **11**(12), 1740–1761 (2017)
5. G. Ferri, A. Tesei, P. Stinco, K.D. LePage, A Bayesian occupancy grid mapping method for the control of passive sonar robotics surveillance networks. In: *OCEANS 2019—Marseille* (2019), pp. 1–9
6. J.P. Feuillet, W.S. Allensworth, B.K. Newhall, Nonambiguous beamforming for a high resolution twin-line array. *J. Acoust. Soc. Am.* **97**(5), 3292 (1995)
7. T. Warhonowicz, H. Schmidt-Schierhorn, H. Hstermann, Port/Starboard discrimination performance by a twin line array for a LFAS sonar system, in *Proceedings of Underwater Defense Technology (UDT)*, Europe (1999), p. 398
8. A. Poulsen, D. Eickstedt, J. Ianniello, Bearing stabilization and tracking for an AUV with an acoustic line array, in *Proceedings of the IEEE/MTS Oceans 2006 Conference*, Boston, MA, USA
9. F.J. Fahy, *Sound Intensity*, 2nd edn. (E and FN Spon, London, 1995)
10. D.L. Hutt, P.C. Hines, A.A.J. Hamilton, Measurements of underwater sound intensity vector, Oceans '99. in *MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings (IEEE Cat. No.99CH37008)*, Seattle, WA, USA, vol. 2 (1999), pp. 717–722
11. R. Hickling, W. Wei, Finding the direction of a sound source using a vector sound intensity probe. *J. Acoust. Soc. Am.* **94**(4), 2408–2412 (1993)
12. M.J. Berliner, J.F. Lindberg, Acoustic particle velocity sensors: design, performance and applications, in *AIP Conference Proceedings*, vol. 368 (1995)
13. J.N. Maksym, M.S. Sandys-Wunsch, Adaptive beamforming against reverberation for a three-sensor array. *J. Acoust. Soc. Am.* **102**(6), 3433–3438 (1997)
14. J. Gebbie, M. Siderius, P.L. Nielsen, J.H. Miller, S. Crocker, J. Giard, Small boat localization using adaptive three-dimensional beamforming on a tetrahedral and vertical line array, in *Proceedings of Meetings on Acoustics ICA2013* (Acoustical Society of America, 2013), p. 070072
15. G.L. D'Spain, J.C. Luby, G.R. Wilson, R.A. Gramann, Vector sensors and vector sensor line arrays: comments on optimal array gain and detection. *J. Acoust. Soc. Am.* **120**, 171–185 (2006)
16. A. Nehoray, E. Paldi, Acoustic vector-sensor array processing. *IEEE Trans. Signal Process.* **42**, 9 (1994)
17. J. Cao, J. Liu, J. Wang, X. Lai, Acoustic vector sensor: reviews and future perspectives. *IET Signal Proc.* **11**, 1–9 (2017)
18. P. Stinco, A. Tesei, G. Ferri, S. Biagini, M. Micheli, B. Garau, K.D. LePage, L. Troiano, A. Grati, P. Guerrini, Passive acoustic signal processing at low frequency with a 3-d acoustic vector sensor hosted on a buoyancy glider. *IEEE J. Ocean. Eng.* **46**, 283–293 (2020)
19. Y. Doisy, L. Deruaz, S.P. Van Ijsselmuiden, S.P. Beerens, R. Been, Reverberation suppression using wideband Doppler-sensitive pulses. *IEEE J. Ocean. Eng.* **33**(4), 419–433 (2008)
20. J. Groen, S.P. Beerens, R. Been, Y. Doisy, E. Noutary, Adaptive port-starboard beamforming of triplet sonar arrays. *IEEE J. Ocean. Eng.* **30**(2), 348–359 (2005)
21. G.W.M. Van Mierlo, S.P. Beerens, R. Been, Y. Doisy, E. Trouw, Port-starboard discrimination on hydrophone triplets in active and passive towed arrays, in *Proceedings of Underwater Defense Technology (UDT)*, Hamburg, Germany (1997), p. 176181



22. S.P. Beerens, R. Been, J. Groen, Y. Doisy, E. Noutary, PortStarboard discrimination on hydrophone triplets in active and passive towed arrays, in *Proceedings of Underwater Defense Technology (UDT), Pacific* (2000), p. 6368
23. P. Stinco, A. Tesei, A. Maguer, F. Ferraioli, V. Latini, L. Pesa, Sub-bands beam-space adaptive beamformer for port-starboard rejection in triplet sonar arrays, in *2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018—Proceedings*
24. D.P. Williams, Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks, in *23rd International Conference on Pattern Recognition (ICPR)* (2016)
25. G. De Magistris, P. Stinco, J.R. Bates, J.M. Topple, G. Canepa, G. Ferri, A. Tesei, K.D. LePage, Automatic object classification for low-frequency active sonar using convolutional neural networks, in *OCEANS 2019 MTS/IEEE SEATTLE*
26. G. De Magistris, M. Uney, P. Stinco, G. Ferri, A. Tesei, K. Le Page K, Selective information transmission using convolutional neural networks for cooperative underwater surveillance, in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)* (2020), pp. 1–8
27. P. Stinco, G. De Magistris, A. Tesei, K.D. LePage, Automatic object classification with active sonar using unsupervised anomaly detection, in *2020 28th European Signal Processing Conference (EUSIPCO)* (2021), pp. 46–50
28. C.W. Ten, J. Hong, C.C. Liu, Anomaly detection for cybersecurity of the substations. *IEEE Trans. Smart Grid* **2**(4), 865–873 (2011)
29. M. Ahmed, A.N. Mahmood, M.R. Islam, A survey of anomaly detection techniques in financial domain. *Future Gener. Comput. Syst.* **55**, 278–288 (2016)
30. Y. Chen, N.M. Nasrabadi, T.D. Tran, Sparse representation for target detection in hyperspectral imagery. *IEEE J. Sel. Top. Signal Process.* **5**, 629–640 (2011)
31. C. Spence, L. Parra, P. Sajda, Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model, in *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*
32. G. Mishne, I. Cohen, Multiscale anomaly detection using diffusion maps. *IEEE J. Sel. Top. Signal Process.* **7**, 111–123 (2012)
33. J. McKay, V. Monga, R.G. Raj, Robust sonar ATR through Bayesian pose-corrected sparse classification. *IEEE Trans. Geosci. Remote Sens.* **55**, 5563–5576 (2017)
34. V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey. *ACM Comput. Surv. (CSUR)* **41**(3), 1–58 (2009)
35. D.M. Hawkins, *Identification of Outliers*, vol. 11 (Chapman and Hall, London, 1980)
36. D.E. Denning, An intrusion-detection model. *IEEE Trans. Softw. Eng.* **2**, 222–232 (1987)
37. M. Micheli, A. Tesei, G. Ferri, P. Stinco, Adaptive filter of seabed clutter onboard the AUVs of an active multistatic sonar network, in *2018 OCEANS—MTS/IEEE Kobe Techno-Oceans (OTO)*
38. P. Gandhi, S. Kassam, Analysis of CFAR processors in nonhomogeneous background. *IEEE Trans. Aerosp. Electron. Syst.* **24**(4), 427–445 (1988)
39. R.M. Haralick, L.G. Shapiro, *Computer and Robot Vision*, vol. I (Addison-Wesley, Reading, 1992), pp.28–48

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)