2022 2nd International Conference on Intelligent Technologies (CONIT)
Karnataka, India. June 24-26, 2022

Supervised Learning Techniques for Detection of Underwater Acoustic Sources

Sanjana K. R. Prasad
Department of Electronics and Communication Engineering
PES University
Bengaluru 560085, India
sanjanaprasadkr@gmail.com

Sanjeev Gurugopinath
Department of Electronics and Communication Engineering
PES University
Bengaluru 560085, India
sanjeevg@pes.edu

2022 International Conference on Intelligent Technologies (CONIT) | 978-1-6654-8407-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/CONIT55038.2022.9848188

***Abstract*—Detection of acoustic sources in an ocean is an important and a challenging problem. In this work, we present a performance comparison of supervised learning techniques for detection of an acoustic source. Performance evaluation of the considered classifiers, which include support vector machine (SVM), logistic regression, decision trees, random forest and naive Bayes, are carried out on synthetic and experimental datasets. The noise model is assumed to follow a generalized Gaussian distribution with heavy tails, which depicts a practical scenario. Our extensive set of experiments indicate that SVM outperforms all other classifiers, in terms of detection accuracy.**

***Index Terms*—Classification, detection, supervised learning, support vector machine, underwater acoustics.**

I. Introduction

Underwater target detection is of prime interest in acoustic signal processing, which is the fundamental problem to be tackled in the development of marine resources and ocean floor exploration, and is of strategic importance in military applications. The complexity of the ocean and its low signal-to-noise ratio (SNR) environment makes target detection challenging. The SNR gives the measure of the strength of the signal from a target, relative to the background underwater noise. The radiated signal from the target, merged with the ambient noise in ocean makes classification task harder [1].

Several signal processing techniques have been employed for detection, including matched filtering [2]. However, implementation of these techniques require full / partial knowledge of the ocean environment and ambient noise statistics, which may be difficult to obtain in practice [3]. On the other hand, blind techniques such as energy detection do not yield a good performance under low SNR. Towards this end, machine learning (ML) techniques have been employed more recently [4]. Several ML techniques have been employed for target detection for both active and passive SONAR applications [5], [6]. Passive systems simply receive the signal emitted from the target, while active sonar systems use transducers to emit an acoustic signal into the ocean and listen for an echo from the target. Compared to conventional techniques, the ML-based techniques have been proven to be capable of achieving good accuracy with complex patterns and models [7]. Deep learning (DL), a subset of ML, has been found to have commendable accuracy in classification problems, and has the potential to automatically extract features which effectively represent passive sonar target signatures in the variable realms of the ocean [8]–[11]. Target recognition based on ship-radiated noise are also a part of SONAR, where DL algorithms have potential to enhance detection accuracy [12]–[14]. However, DL algorithms require large amount of training, and have poor interpretability.

In this paper, we carry out a performance comparison of some well-known supervised learning (SL) algorithms for detection of underwater acoustic sources. In particular, we consider the following algorithms, namely, support vector machine (SVM), logistic regression, decision trees, random forest and naive Bayes classifiers. We evaluate the performance of these SL algorithms on both synthetic and real-world experimental datasets. The synthetic dataset is generated by following the normal mode theory of oceans, while the experimetal dataset follows the experimentally captured dataset downloaded from the historic naval sound and video database website [8]. Noise is synthetically added to both datasets, following a single parameter generalized Gaussian distribution, which is of practical relevance [2]. Our extensive experimental results show that the SVM outperforms the other SL algorithms consistently, and thereby is a good candidate for real-world applications.

II. Data Model and Problem Formulation

A waveguide-driven model is considered for the ocean, and the water layer between the depths $0 \leq z \leq D$ is assumed to have a sound speed profile $c(z)$ and density $\rho(z)$. In the far field region of an acoustic source at an unknown location, we consider a uniform linear array of N sensors with an inter-spacing d [2]. The number of snapshots of the signal at the

978-1-6654-8407-7/22/$31.00 ©2022 IEEE

n^{th} sensor, for $n = 1, \ldots, N$ is assumed to be M. Let the received signal vector at the array (without noise) be denoted by $\mathbf{x}(m)$, $m = 1, \ldots, M$, where

$$\mathbf{x}(m) \triangleq [x(1), x(2), \ldots, x(M)]^T, \tag{1}$$

for $m = 1, \ldots, M$. Here, $[\cdot]^T$ denotes transpose of a matrix.

A. Signal Model

Following the normal mode theory, we write [15], [16],

$$\mathbf{x}(m) = \mathbf{A}\mathbf{z}(m), \tag{2}$$

where

$$\mathbf{A} = [\mathbf{a}_1(\theta) \cdots \mathbf{a}_K(\theta)], \tag{3}$$

$$\mathbf{a}_k(\theta) = \left[1 \quad e^{i\kappa_k d \cos\theta} \quad \cdots \quad e^{i(N-1)\kappa_k d \cos\theta}\right], \tag{4}$$

and $\mathbf{z}_k$ is a $1 \times K$ column vector whose k^{th} element is

$$\mathbf{z}(m)_k = \frac{\psi_k(z_a)\psi_k(z)e^{i\zeta_k - \alpha_k r}}{\sqrt{\rho_k r}}, \tag{5}$$

with $\psi_k(z)$, α_k and ζ_k respectively denote the eigenfunction, attenuation coefficient and wavenumber corresponding to the k^{th} normal mode out of a total of K modes at frequency f. We consider a cylindrical coordinate system (r, z, θ) which denotes a position in the ocean.

B. Noise Model

Typically, it is assumed that the noise is Gaussian distributed with a known/unknown variance, independent and identically distributed across $m = 1, \ldots, M$. However, the probability density of the ambient noise in the ocean is known to have heavier tails. In this work, we assume that the noise vector follows a generalized Gaussian distribution (GGD) with zero mean and variance σ^2. GGD is of large service for synthetic modeling in signal processing [17]. The probability density of a noise sample W is given by

$$f_N(x) = \frac{\beta}{2\beta\Gamma(\frac{1}{\beta})} \exp\left(\frac{-|x|^\beta}{\alpha}\right), \tag{6}$$

where the parameter $0 < \beta \leq 2$ dictates the probability density tail [18]. It is well-known that the Laplace and Gaussian distributions are special cases of GGD, with $\beta = 1$ and $\beta = 2$, respectively.

C. Hypothesis Testing Problem

The problem of acoustic signal source detection in an ocean is formulated as the hypothesis testing problem:

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{y}(m) = \mathbf{w}(m), \\ \mathcal{H}_1 &: \mathbf{y}(m) = \mathbf{x}(m) + \mathbf{w}(m), \end{aligned} \tag{7}$$

for $m = 1, \ldots, M$, where $\mathcal{H}_0$ and $\mathcal{H}_1$ denote the noise-only and signal-present hypotheses, respectively, and $\mathbf{w}(m)$ is the noise vector, whose entries are assumed to be independent and identically distributed GGD random variable with parameter β.

In the next section, we present a background on the different supervised learning algorithms that we have considered in this work.

III. Supervised Learning Algorithms

The central goal in supervised learning is to predict an output Y, given the input x. This is obtained by first training the ML algorithm with sample data and labels, followed by testing. For a model with a fixed number of categorical classes, the task of prediction of the class of Y is classed as classification. In the classification problem, supervised learning yields an estimator to predict the class of an instance, dictated by a set of features. These features are given as input to the learning technique, which then compares its output to the actual outputs given by the training set, and performs optimization to minimize the error [19]. A generic model defined with an unknown set of parameters is given by

$$Y = G(x|\theta) \tag{8}$$

where $G(\cdot)$ is the discriminant function, θ is the set of its parameters, and Y is the class label in the case of classification. In this work, we have used the following learning algorithms for the classification of acoustic signals from an underwater source in a low SNR environment [20].

A. Logistic Regression

Logistic regression is a machine learning classification approach that employs one or more independent variables to predict a result using statistical approaches. In a binary logistic regression model, the outcome is measured using a dichotomous variable, i.e., there are only two possible outcomes. Logistic regression identifies the best fit between a dependent variable and a set of independent variables. The mean of a distribution over a binary variable must always lie between 0 and 1. The logistic sigmoid function

$$y = \frac{1}{1 + e^{-x}}, \tag{9}$$

is used so the output of the linear function is in the interval $(0, 1)$, and that value is interpreted as a probability

$$p(y = 1|x; \theta) = \sigma(\theta < x). \tag{10}$$

In general, there is no closed-form solution for its optimal weights. Instead of trying to achieve optimization of weights as in linear regression, we seek to maximize the log-likelihood. This can be achieved by employing gradient descent to minimize the negative log-likelihood. Logistic

regression incorporates commendable probabilistic interpretation unlike Decision Tress or support vector machines.

B. Support Vector Machine

Similar to the logistic regression, a support vector machine (SVM) is driven by a linear function $wx+b$, with its outputs being class identities. The goal here is to find a hyperplane in an N-dimensional space that successfully classifies the data points (where N is the number of features), with the maximization of the minimum distance from the hyperplane to the nearest instance. Based on the Vapnik-Chervonenkis theory, it is particularly robust to high dimensional data. SVMs use the powerful *kernel trick*, to deal with linearly inseparable data. It projects nonlinear combinations of the original features, onto a higher-dimensional space via a mapping function. Predictions can be made using the function

$$f(x) = c + \sum_t \gamma_i k(\mathbf{x}, \mathbf{x^{(t)}}), \tag{11}$$

where $\mathbf{x^t}$ are training instances, γ is a vector of coefficients and $k(\mathbf{x}, \mathbf{x^t}) = \psi \mathbf{x^T} \psi(\mathbf{x^{(t)}})$ is the kernel where $\psi(\mathbf{x})$ denotes feature function. This function is non-linear as a function of x, but linear in the space mapped by ψ. The *Gaussian Kernel* is most commonly used.

C. Naive Bayes

Naive Bayes assumes class conditional independence, which means that there is a strong assumption of independence amongst child nodes. It posits that the existence of one feature in a class is unrelated to the presence of any other feature. This algorithm is a Bayesian network, which is a probabilistic graphical model for representing variables and their conditional dependencies, which induces a distribution:

$$P(c, x_t ... x_n) = P(c) \prod_{t=1}^{n}, \tag{12}$$

where c is class label, x_t is the value of an attribute X_t, $P(C)$ is class prior and $P(X_t|C)$ are conditional distributions. Once we have learned a Naïve Bayes classifier from data, the class label c^* which has maximum posterior probability (also known as the maximum a posteriori (MAP) class) can be labelled for new instances as follows:

$$c^* = argmax P(c|x_t, ...x_n). \tag{13}$$

The Naive Bayes model is simple to construct and is especially good for huge data sets. Naive Bayes is known to outperform several advanced classification techniques.

D. Decision Tree

A decision tree is an easy to interpret, efficient hierarchical data structure. The procedure for determining the class of a given dataset in a decision tree starts at the root node of the tree. Each node in this algorithm checks the values of the root attribute with the values of the record attribute from the actual dataset, and then follows the branch to the next node based on the comparison. This process is repeated until the tree's leaf node is reached. As the tree grows, so do the number of nodes, which poses a problem in making statistically strong decisions about class representation. This problem, termed the *data fragmentation problem*, can be avoiding by declaring a threshold below which further splitting is forbidden. Due to its non-parametric nature, outliers do not significantly affect this model. Decision trees have good generalization ability and provide relatively high performance for less computational power.

E. Random Forest Classifier

A random forest is made up of a number of decision trees that work together to form an ensemble. A random forest consists of a collection of tree structured classifiers

$$h(x, \phi_m), m = 1, \ldots, \tag{14}$$

where ϕ_m are independent and identically distributed random vectors. Each tree in the random forest produces a class prediction from the input x, and the class with the most votes becomes the prediction of our model. They are fast and easy to interpret and visualize, however the algorithm slows down for real-time prediction with increasing number of trees. Random Forests utilize the statistical technique of Bootstrapping, to address the problem of overfitting due to high variance in decision trees. This technique draws several random samples with replacement from the dataset, which enables the creation of multiple and smaller random datasets from the same distribution.

IV. Performance Metrics

Good accuracy in classification is the primary concern. Accuracy is a commonly used metric, which is defined as the percentage of correct predictions the algorithm achieves for the given test data. The outcome of a classification is true when the prediction is correct, and false when there is a mismatch between the predicted and actual class labels. The label can be predicted as positive or negative. There are four types of outcomes that could occur.

1) True Positive (TP): Number of times the model correctly classifies a positive sample as positive.
2) True Negative (TN): Number of times the model correctly classifies a negative sample as negative.

3) False Positive (FP): Number of times the model incorrectly classifies a negative sample as positive.
4) False Negative (FN): Number of time the model incorrectly classifies a positive sample as negative.

Accuracy is defined as the ratio between the number of correct predictions to the total number of predictions, i.e.,

$$Accuracy = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{15}$$

The probability of detection of an underwater acoustic source is calculated as the ratio between the number of positive samples correctly classified as positive, to the total number of positive samples. It is given by:

$$P_d = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{16}$$

V. Datasets

A. Model-based Synthetic Dataset

To establish a baseline, we construct a synthetic dataset for acoustic signals following the normal mode theory discussed in Sec. II-A, with parameters chosen as follows. Speed of sound in water is 1500 m/s, speed of sound at bottom of the ocean 1700 m/s, ratio of densities at bottom of the ocean and water is 1.5, depth of ocean is 150 m and attenuation at bottom of the ocean is 0.2 dB/wavelength. The Kraken normal mode program has been referred to for computation of the eigen functions $\psi_k(z)$, wavenumbers α_k and attenuation coefficients ζ_k [21]. The noise samples are generated from a GGN distribution with different values of β, namely $0.5, 1$ and 2.

B. Experimental Datasets

Real-world datasets for training of the models have been obtained from the publicly available Historic Naval Sound and Video database website [8]. The website contains recordings of underwater targets from WWII, including torpedo, mine-sweeper, and foxer sounds, submarine sounds, marine life sounds and acoustic signals from other underwater targets. For this paper, we have considered three declassified training recordings of USN submarines on an early radio sonobuoy system.

1) Dataset 1: "Submarine moving towards hydrophone".
2) Dataset 2: "Approaching and leaving small submarine with little background noise".
3) Dataset 3: "Small submarine propulsion with little background noise".

The audio files were first converted from MP3 to WAV format. Preprocessing the dataset included resampling the audio signals from 22 kHz to 2 kHz, as preliminary observation revealed that only the low frequency features were of importance. Following this, the fast Fourier transform of the signals were taken, and the signals were normalized. The total number of instances in our dataset is $10,000$, with 8000 of the instances utilized for training, and 2000 for testing.

VI. Results and Discussion

We next study the performance of the considered SL techniques, in terms of detection accuracy. As mentioned earlier, we consider GGN noise with varying values of $\beta = 0.5, 1$ and 2. In particular, we present a detailed performance comparison of the SL algorithms for all the datasets explained in Section V.

A. Model-based Synthetic Dataset

Figure 1 shows the variation of accuracy of the SL techniques with SNR. As expected, accuracy improves with an increase in SNR for all cases. Further, as shown in Fig. 1a, the accuracy values corresponding to $\beta = 0.5$ is highest as compared to other β values, with an accuracy of 73.4% at SNR = 0 dB obtained by SVM, LR and NB algorithms. Accuracy corresponding to $\beta = 2$, which results in Gaussian noise, has the least accuracy as seen from Fig. 1c. The SVM, LR and NB algorithms perform equally well and well-exploit the heavy tails in a GGN distribution as β deviates away from 2. The trends in Fig. 2 supports the observation.

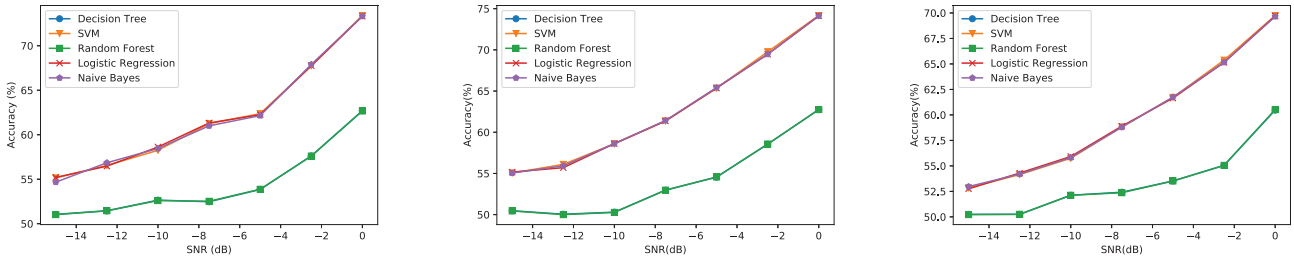
For benchmarking, we also carry out a performance comparison between the theoretical values and simulated values of the probability of detection for SVM, LR and NB algorithms as shown in Table I. The theoretical P_d values correspond to $\beta = 2$, and are calculated for SNR values of -5 dB, 0 dB and 3 dB. The simulated values are noted to be very close to the theoretical values.

B. Experimental Datasets

Classification results on the experimental real-world datasets follow a similar trend as seen by the model-based synthetic dataset. The SL algorithms have been identically trained and tested on the experimental datasets 1, 2 and 3, and the performances are given in Figs. 3, 4 and 5, respectively. It is observed that a smooth variation in accuracy is not obtained in all the SL algorithms, but SVM consistently outperforms all the other algorithms. The performance of LR is close to that of the SVM in several occasions. The RF algorithm performs the least even in this scenario. To summarize, the SVM algorithm works well in a practical scenario, which is established given its superior performance across the experimental datasets.

VII. Conclusion

We carried out a detailed performance comparison of supervised learning techniques for detection of an acoustic source. We considered the generalized Gaussian distribution model for noise, which is of practical relevance. A good



(a) Variation with SNR, with $\beta = 0.5$. (b) Variation with SNR, with $\beta = 1$. (c) Variation with SNR, with $\beta = 2$.

Fig. 1: Performance comparison of the supervised learning algorithms for the dataset generated using Monte Carlo simulations under GGN, for different values of β.

	SNR = -5 dB		SNR = 0 dB		SNR = 3 dB	
Algorithm	P_d, Th.	P_d, Sims.	P_d, Th.	P_d, Sims.	P_d, Th.	P_d, Sims.
SVM	0.60	0.62	0.68	0.70	0.76	0.77
LB	0.60	0.61	0.68	0.70	0.76	0.77
NB	0.60	0.62	0.68	0.70	0.76	0.77

TABLE I: Comparison between theoretical and simulated probability of detection for the dataset generated using Monte Carlo simulations under additive white Gaussian noise ($\beta = 2$).

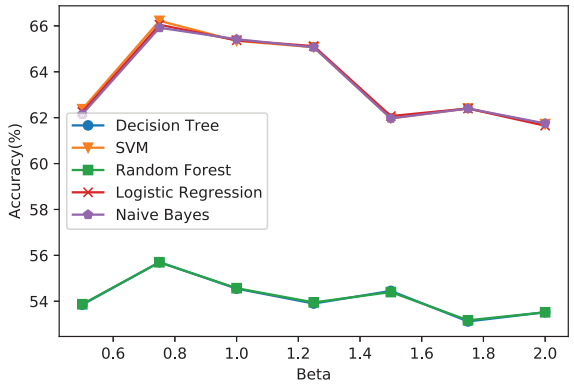


Fig. 2: Performance comparison of the supervised learning algorithms for the dataset generated using Monte Carlo simulations under GGN, for different values of β.

match between the theoretical and simulated probability of detection values exhibited the accuracy of the considered SL algorithms. It was observed that SVM outperformed LR, DF, RF and NB classifiers with consistently good accuracy.

REFERENCES

[1] A. Ganesan, S. M. Santhanam, V. Padmanabhan, S. Vivekanandan, and V. Selvaraj, "Coherence analysis of ambient noises in shallow water for underwater communication," *Journal of Marine Science and Technology (Taiwan)*, vol. 25, pp. 311–318, 01 2017.

[2] P. V. Nagesha, G. V. Anand, N. Kalyanasundaram, and S. Gurugopinath, "Detection, enumeration and localization of underwater acoustic sources," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[3] X. Guan, R. Li, and P. Hu, "A novel detection method for weak underwater acoustic targets based on rusboost," in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, vol. 9, 2020, pp. 2257–2261.

[4] W. Zhang, Y. Wu, Y. Lin, L. Ma, K. Han, Y. Chen, and C. Liu, "Underwater target detection based on machine learning," in *2020 IEEE 3rd International Conference on Information Communication and Signal Processing (ICICSP)*, 2020, pp. 210–214.

[5] H. Yang, S.-H. Byun, K. Lee, Y. Choo, and K. Kim, "Underwater acoustic research trends with machine learning: Active sonar applications," *Journal of Ocean Engineering and Technology, The Korean Society of Ocean Engineers*, vol. 34, no. 4, pp. 277–284, 2020.

[6] H. Yang, K. Lee, Y. Choo, and K. Kim, "Underwater acoustic research trends with machine learning: Passive sonar applications," *Journal of Ocean Engineering and Technology*, vol. 34, 06 2020.

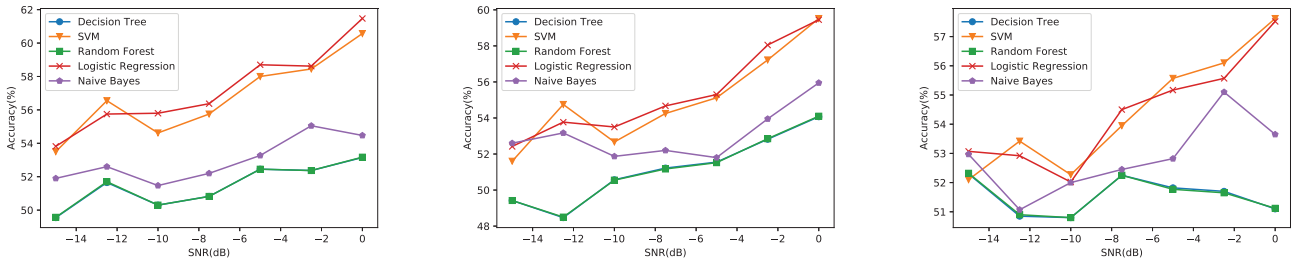
[7] S. Kamal, S. K. Mohammed, P. R. S. Pillai, and M. H. Supriya, "Deep learning architectures for underwater target recognition," in *2013 Ocean Electronics (SYMPOL)*, 2013, pp. 48–54.

[8] H. Yue, L. Zhang, D. Wang, Y. Wang, and Z. Lu, "The classification of underwater acoustic targets based on deep learning methods," in *Proceedings of the 2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI 2017)*. Atlantis Press, 2017/06, pp. 526–529.

[9] H. Yang, G. Xu, S. Yi, and Y. Li, "A new cooperative deep learning method for underwater acoustic target recognition," in *OCEANS 2019 - Marseille*, 2019, pp. 1–4.

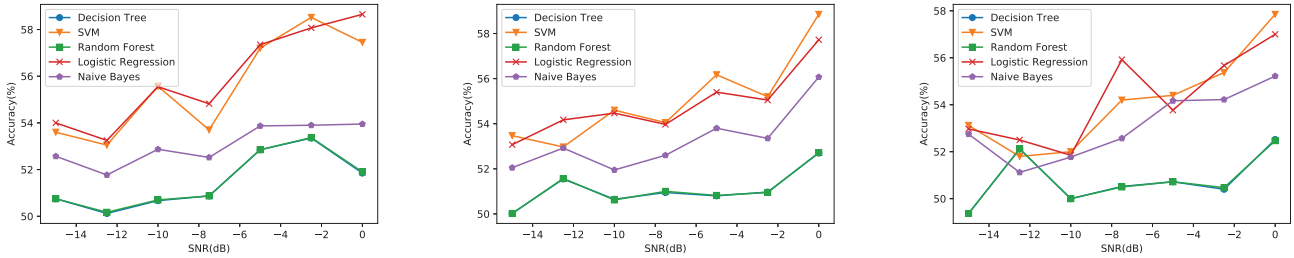
[10] S. Tian, D. Chen, H. Wang, and J. Liu, "Deep convolution stack for waveform in underwater acoustic target recognition," *Scientific Reports*, vol. 11, 05 2021.

[11] J. Chen, J. Liu, C. Liu, J. Zhang, and B. Han, "Underwater target recognition based on multi-decision lofar spectrum enhancement: A deep learning approach," 04 2021.



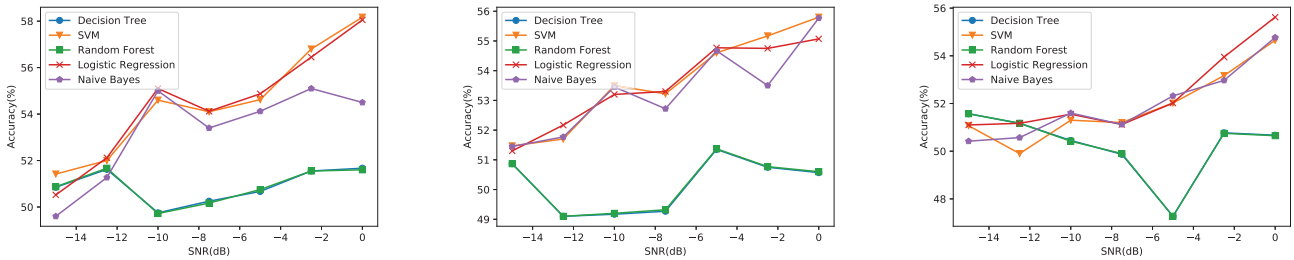
(a) Variation with SNR, with $\beta = 0.5$. (b) Variation with SNR, with $\beta = 1$. (c) Variation with SNR, with $\beta = 2$.

Fig. 3: Performance comparison of the supervised learning algorithms for the Dataset 1 under GGN, for different values of β.



(a) Variation with SNR, with $\beta = 0.5$. (b) Variation with SNR, with $\beta = 1$. (c) Variation with SNR, with $\beta = 2$.

Fig. 4: Performance comparison of the supervised learning algorithms for the Dataset 2 under GGN, for different values of β.



(a) Variation with SNR, with $\beta = 0.5$. (b) Variation with SNR, with $\beta = 1$. (c) Variation with SNR, with $\beta = 2$.

Fig. 5: Performance comparison of the supervised learning algorithms for the Dataset 3 under GGN, for different values of β.

[12] H. Yang, S. Shen, X. Yao, M. Sheng, and C. Wang, "Competitive deep-belief networks for underwater acoustic target recognition," *Sensors*, vol. 18, no. 4, 2018.

[13] H. Yang, J. Li, S. Shen, and G. Xu, "A deep convolutional neural network inspired by auditory perception for underwater acoustic target recognition," *Sensors*, vol. 19, no. 5, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/5/1104

[14] X. Luo, Y. Feng, and M. Zhang, "An underwater acoustic target recognition method based on combined feature with automatic coding and reconstruction," *IEEE Access*, vol. 9, pp. 63 841–63 854, 01 2021.

[15] A. Haug, R. Graves, and H. Überall, "Normal-mode theory of underwater sound propagation from stationary multipole sources: Results for a realistic sound-speed profile," *Acoustical Society of America Journal*, vol. 57, pp. 1052–1061, 05 1975.

[16] B. Jensen, F. A. Kuperman, W. B. Porter, M., and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed. Springer, 2011.

[17] M. Novey, T. Adali, and A. Roy, "A complex generalized gaussian distribution— characterization, generation, and estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1427–1433, 2010.

[18] S. Gurugopinath, R. Muralishankar, and H. N. Shankar, "Differential entropy-driven spectrum sensing under generalized gaussian noise," *IEEE Communications Letters*, vol. 20, no. 7, pp. 1321–1324, 2016.

[19] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51–62, 2017.

[20] A. Singh, N. Thakur, and A. Sharma, "A review of supervised ma-

chine learning algorithms," in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, 2016, pp. 1310–1315.
[21] M. B. Porter, *The KRAKEN normal mode program*, 1992.
[22] J. Liu, I. Nevat, P. Zhang, and G. W. Peters, "Multimodal data fusion in sensor networks via copula processes," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6.