Performance Evaluation of Linear and Multi-linear Subspace Learning Techniques for Object Classification Based on Underwater Acoustics

Mohtashim Baqar and Syed Sajjad Haider Zaidi
National University of Sciences and Technology
Sector H-12, Islamabad, Pakistan
{*mohtashimbq*} , {*sajjadzaidi*} {*@pnec.nust.edu.pk*}

***Abstract*—Underwater object identification based on acoustic sequence is a complex task, mainly, because of the non-stationary nature of the underwater environment. Moreover, the ambient conditions contribute heavily to varying temporal and spectral characteristics of the source. Further, the characteristic features of a source lie within its spectrum whereas pure spectral contents are more robust to variations along the time and frequency axis. In this work, performance of two different class of learning approaches i.e. linear and multi-linear subspace learning, have been evaluated. Moreover, spectral features are used as inputs to both the said approaches. Two linear subspace learning techniques, namely, principal component analysis (PCA) and linear discriminant analysis (LDA) along with one multi-linear subspace learning (MSL) technique, namely, multi-linear principal component analysis (MPCA) have been used. Performance of the system was evaluated using two sets of data i.e. raw acoustic dataset having samples belonging to 4 distinct classes of ships and a synthetic dataset downloaded from DOSITS, having acoustic samples belonging to 20 distinct classes of underwater objects i.e. sea species and man-made objects. For raw acoustic database, ships signatures were collected in the Indian ocean. Further, two-pass split window (TPSW) method was used to remove background noise from the processed raw acoustic samples. For classification, two neural classifiers were used, namely, robust variable learning rate feed-forward neural network (RVLR-NN) and convolution neural network (CNN). All simulations have been conducted in MATLAB. Further, the system was evaluated under the effect of noise i.e. additive white Gaussian noise (AWGN) at different levels of signal-to-noise ratio (SNR). In addition, dimensions of the feature set were also varied and effects of dimensionality reduction on classification accuracies were observed. Simulation results observed have shown that the combination of MPCA-CNN produced best classification results with an accuracy of up to 99.4%.**

***Index Terms*—Acoustic signal processing, computational complexity, dimensionality reduction, machine learning, pattern analysis, sound navigation and ranging (sonar) detection, underwater acoustics.**

I. Introduction

Detection, classification and tracking of objects based on underwater acoustics has been a growing area of research, primarily because it has a lot of applications in military and commercial settings. The need of improvement in port security and the security of coastal & offshore operations has led to majority of the research and development being made in the area of underwater acoustics. In addition, the recent awareness and care for endangered marine wildlife has also resulted in a lot of data collection and conduction of studies [1]. Moreover, for marine object detection and recognition sonar signal processing systems have long been in use since the turn of the 20^{th} century, although major contributions to the field were made post world war II [2]. Further, last 60 years have seen a lot of contributions being made to the said area of work with majority of the studies being conducted by the defence organizations and due to the sensitive nature of the issue, not much of the work has been made public [3]. According to a survey by Lampert [4], it is difficult to exhibit a direct comparison between the studies available in literature as there exist large variations in the way results have been presented i.e. variations in testing conditions, choice of data set - synthetic or real-time, type of analysis - qualitative or quantitative and variations in operating conditions of the sonar platform or target etc.

Also, detection, classification and tracking of objects based on acoustic sequences in underwater environment has been a complex task due to the presence of noise, interfering clutter of sources and most importantly due to the non-stationary nature of the underwater environment [5] [2]. Signals acquired from the sonar platform comprised of noise radiated from underwater vessels or from marine species i.e. mantis shrimp or whales etc. Every signal has it's own characteristics, which are labelled by a sonar operator either by listening or by investigating the spectrograms of the processed signals [6]. However, this is not an objective approach, so this raises the need for an expert system to perform automated classification. Further, characteristic features of an underwater source lie within it's spectral content. Spectral estimates are more robust compared to parametric and statistical models, however, systems making use of pure spectral content for detection and classification of underwater objects met with little success in early era of sonar signal processing [7]. In a work [3], authors presented an experimental study by extracting ship signatures in an urban harbour and used detection envelope modulation on noise (DEMON), an spectral estimation technique, and a proposed modified version of DEMON to discriminate ships in a cluttered environment. In another work [8], authors used spectral estimates and neural networks to detect and classify underwater objects, respectively. Another study [9] made use of spectral estimates to classify the ongoing processes in underwater environment via an autonomous vehicle. Similarly in articles [10]–[13], authors have used spectral estimates for detection of underwater objects. Spectral estimates are usually in form of a vector comprising magnitude of energy at specific frequency bins. Spectral estimates with increased resolution may help the system to classify targets better, so to increase resolution of the feature map, number of frequency bins are to be increased which results in greater size of the feature vector while characteristic features of a source doesn't cover the entire spectrum and are usually present in form of tonal components in the spectrum [14]. Further, characteristics of a source are sparsely populated in the feature map and with larger size of the feature vector, the learning process will get complex for the classifier, which may result in poor performance levels of the detection and classification system. Moreover to cope with the problem, many researchers have applied feature learning techniques to the spectral estimates i.e. linear sub-space learning approaches. Among many, principal component analysis (PCA) has been the most used. A recent survey indicated that PCA is still one of the more sought after learning and dimensionality reduction mechanism used in almost every discipline requiring data mining and analysis [15]. In [16], authors used compressed wavelet feature set via principal component analysis (PCA) to classify objects based on underwater transients. Further, object classification was made using a feed-forward neural network. In [17], principal component analysis was applied to reduce the dimensionality of the extracted feature set for classifying marine objects based on underwater transients. Artificial neural networks were used for performing object classification. Similarly in articles [17]–[20], principal component analysis was used for feature learning and dimensionality reduction to help in classification of underwater processes and objects.

Further, for classification of underwater objects, a lot of the available studies in literature have used artificial neural networks and it's variants. Primarily, this is because of their ability to learn and cluster [5], [21]. In a study [22], authors have used neural network classifiers in identification of models of the underwater vehicles. Studies in [23]–[26], also made use of artificial neural networks for underwater object classification.

In this work, performance evaluation of linear and multi-linear subspace learning approaches have been made for an underwater detection and classification system. For the said purpose, two datasets have been used i.e. a synthetic dataset [27], [28], having samples belonging to 20 distinct classes of underwater objects and a raw sonar dataset, having samples belonging to 4 distinct classes of ships.

Ship signatures for the raw sonar dataset were collected in the Indian ocean. Further, spectral estimates have been used as inputs to both the said learning approaches. Moreover, for spectral estimation low frequency analysis and ranging (LOFAR) method has been used. Also, the spectral estimates were processed to remove any noise present in the spectrum i.e. white noise, using the two-pass split window (TPSW) method. Further, for comparison and evaluation, two widely used linear subspace learning techniques i.e. principal component analysis (PCA) and linear discriminant analysis (LDA), and one relatively new multi-linear subspace learning approach i.e. multi-linear principal component analysis [29], have been used. For object classification, two renowned classifiers have been employed in this study, a feed-forward neural network with modified back-propagation algorithm [30], [31] i.e. robust variable learning rate feed-forward neural network (RVLR-NN), and convolution neural network (CNN), a deep learning approach.

In recent times, deep learning and multi-linear subspace learning approaches have shown to outperform many standard and conventional methods in many research applications related to image, speech and acoustics [32]–[34] but there haven't been many studies available investigating their performance in underwater applications. Also, CNN and MPCA have both been applied successfully to problems where conventional techniques have failed to perform. Moreover, both techniques have worked well in environments that uses tensor objects as inputs i.e 2D or 3D matrices. In addition, both techniques have performed well and given sustained performance levels where size, dimension and complexity of data is an issue. Further, effects of dimensionality reduction have also been observed in presence of noise at different levels of signal-to-noise ratio (SNR) i.e. additive white Gaussian noise (AWGN). Many studies have shown spectral density of the ambient noise in underwater environment to be relatively flat in the acoustic region i.e. Gaussian. However, some studies have shown it to be non-Gaussian in nature [35].

Next section gives detailed insight into the methodology adopted to model the proposed system.

II. Methodology

A sonar system has four basic building blocks i.e. beamforming for signal acquisition, preprocessing for signal conditioning, detector for extracting discriminating feature set and a classifier for automated classification of objects. In this work, LOFAR + Sub-space learning approaches have been used for feature learning and RVLR-NN & CNN approaches have been used for classification. Brief details of the aforementioned are in the following.

A. Low Frequency Analysis and Ranging (LOFAR)

Low frequency analysis & ranging (LOFAR) [36] is a broadband signal analysis technique and furnishes machinery characteristics of the objects i.e ships and vessels. Moreover, it provides detailed knowledge of machinery noise of the target to the sonar operator. Figure 1 depicts block diagram implementing LOFAR.

Further, on estimation of direction of arrival (DOA), the captured signal is chopped into pieces of short duration referred to as frames. Later, ensembles are passed through a window, Hann window, to select the range of interest. Filtered ensembles are frequency transformed using Fourier transform (FFT). Then, DC component is removed from the resulting spectra and is normalized using TPSW method to remove presence of any bias due to background noise. Spectral estimation for LOFAR has been made using Welch method [37]. It is an improved form of the standard peridiogram estimation techniques. It avoids noisy components getting into the spectra in exchange for good frequency resolution. Moreover, as per Welch method overlapping between two consecutive frames is kept to 50%.

Fig. 1: LOFAR Analysis extracts from consecutive blocks of acoustic sequence. In this particular case, consecutive 25 msec-length blocks were filtered by a Hanning window and retain only 512 bins of frequency. After the full process, each block generates a frequency range from DC to 2200 Hz

1) Two-Pass Split Window (TPSW): Two-pass split window [14] has been applied for spectrum smoothing against background noise. Frequency content of the radiated signal consist of two major components, a broadband component which has a continuous spectrum i.e. noise, and a tonal component, which has a discrete spectra. Mainly, tonal components in the spectra comprises the characteristic features of the source. So, extracting these tonal components is key to having good detection and classification results. Further, TPSW works well in extracting the tonal components from the continuous spectrum. Mathematically, it is defined as,

1) For a signal $f(x)$, a window is selected, centred on the current sample, k, in time as, $R_k = \{k-M, k-M+1, \ldots\ldots, k, \ldots\ldots, k+M-1, k+M\}$. Where, length of the selected window is $2M+1$.
2) In pass-1, mean for the window centred at k is calculated. It is done for all values of k,

$$\hat{f(k)} = \frac{1}{2M+1}\sum_{i=K-M}^{i=K+M} f(i)$$

3) Further, a clipped sequence, $g(k)$, is formed in order to avoid biasing of the estimates of local mean due to the presence of tonal components,

$$g(k) = \begin{cases} \hat{f(k)} & if, \quad f(k) \le \alpha \hat{f(k)} \\ f(k) & if, \quad f(k) > \alpha \hat{f(k)} \end{cases}$$

where, α is a constant. Its typical value is 0.9.
4) Next is pass-2, continuous spectra is again attained by evaluating the local mean of the sequence obtained in pass-1, that is, $g(k)$,

$$\hat{m(k)} = \frac{1}{2M+1}\sum_{i=K-M}^{i=K+M} g(i)$$

On estimation of broadband components, narrow band or tonal components can be evaluated by removing the estimated spectrum from the spectrum of the original signal, i.e.,

$$h(k) = f(k) - \hat{m(k)} \tag{1}$$

The tonal components thus extracted are normalized to avoid any amplitude discrepancies and also, because we are only interested in the patterns present in the spectrum,

$$X = \frac{h}{\parallel h \parallel} \tag{2}$$

B. Subspace Learning Techniques

For feature learning and dimensionality reduction, two linear subspace learning and one multi-linear subspace learning approaches have been used i.e. principal component analysis (PCA), linear discriminant analysis (LDA) and multi-linear principal component analysis (MPCA). Their brief details are in the following.

1) Principal Component Analysis (PCA): It is a statistical data analysis technique [38], also known as an un-supervised classifier and known for increasing the variance amongst the data elements of the input set. It has greater applications in fields of pattern recognition and cryptography. It is very commonly used in applications where dimension of the data set is to be reduced. It transforms the data set in high-dimensional space into a low-dimensional space reducing the overall processing cost with out losing much information. It is used for finding similarities and patterns present in data. Since, there is less chance of efficiently analysing data of higher dimensions analytically, that's where PCA helps the most. It has also been widely used for feature learning in underwater acoustics as explained in previous sections.

2) Linear Discriminant Analysis (LDA): Also a statistical data analysis technique, also known as a supervised classifier. It transforms data elements into an Eigen space while maximizing the projected variance between data elements belonging to different classes and reducing the intra-class projected variance i.e. minimizing the variance between data elements belonging to same class [39]. Further, the key difference between PCA and LDA is that the subsequent deals with labelled data. Objective function for LDA can be mathematically expressed as,

$$\max_{\mathbf{w}} J(w) = \frac{\mathbf{w}^T B_w}{\mathbf{w}^T B_s} \tag{3}$$

where, B_w and B_s represents the between and within class scatter matrices, respectively. Main goal of LDA is to find a vector $\mathbf{w}$ that maximizes the objective function.

3) Multi-linear Principal Component Analysis (MPCA): It is a multi-linear subspace (MSL) learning approach, used for feature learning and dimensionality reduction of the tensor objects i.e. multi-dimensional objects. It is designed to work with tensors of any order i.e. 1D, 2D, 3D etc. Moreover, the objective of MPCA is to find Eigen tensors that tends to capture maximum variations present in the input data. Traditionally, linear subspace learning (LSL) mechanisms i.e PCA and LDA, were used a lot and they have mostly been used to reduce the dimension of the data set. Moreover, almost all LSL methods represent input as vector and solve for transforming the input vectors into an optimal lower-dimensional space. However, with multi-dimensional data LSL methods haven't been effective because it tends to break the natural structure of the objects which results in loss of information and therefore the transformed data usually tends to lose variations present in the original tensors. Moreover, the author in [29] has discussed all mathematical and algorithmic details of multi-linear principal component analysis (MPCA) related to it's implementation. Also, results presented in [29] signifies the importance and need of usage of the MSL approaches for feature learning and dimensionality reduction of tensor objects.

C. Classifier

For automated classification of underwater objects, two classifiers have been used in this study i.e. robust variable learning-rate neural network (RVLR-NN) and convolutional neural network (CNN). Brief details of both are in the following.

1) Robust Variable Learning Rate Feed-forward Neural Network: A supervised multi-layer feed-forward neural network with a modified back-propagation algorithm for weights correction and updation. Weights associated with i^{th} neuron of the $(l-1)^{th}$ layer to the j^{th} neuron in the l^{th} layer are updated as below:

$$\underbrace{w_{ji}^{(l)}(n+1)}_{\text{New Weights}} = \underbrace{w_{ji}^{(l)}(n)}_{\text{Old Weights}} + \mu(n) \underbrace{\delta_j^{(l)}(n)}_{\text{Local Gradient}} \; y_i^{(l-1)}(n) \tag{4}$$

where, $\mu(n)$ represents the time-varying learning rate. It is updated according to the mechanism of robust variable step-size least mean square algorithm [30] as,

$$\mu(n+1) = \alpha\mu(n) + \gamma\rho^2(n) \tag{5}$$

with

$$\rho(n) = \beta\rho(n-1) + (1-\beta)e(n)e(n-1) \tag{6}$$

Such that $0 < \alpha < 1$, $0 < \beta < 1$ and $\gamma > 0$. $\delta_j(n)$ in equation 4 represents the local gradient associated with the j^{th} neuron of the l^{th} layer and can be calculated as follows,

$$\delta_j^{(l)}(n) = \begin{cases} (d_j(n) - y_j^{(L)}(n))\phi_j\prime(v_j^{(L)}(n)) & , OutputLayer \\ \phi_j\prime(v_j^{(L)}(n)) \sum\limits_{\forall K} \delta_k^{(l+1)}(n) w_{kj}^{(l+1)}(n) & , HiddenLayer \end{cases} \tag{7}$$

2) Convolutional Neural Network (CNN): It is a deep learning approach, which has worked well in classification problems where data is tensor, complex and separability is difficult [40]. Convolutional neural networks (CNNs) are inspired with visual mechanism in living beings. The visual cortex in human brain contains a lot of cells that are sensitive to light and detects light in small overlapping sub-regions in the visual fields known as the receptive fields. Complex cells have larger visual fields. Moreover, these cells act as filters in realizing the convolution operation. Convolutional neural network comprised of two major types of layers, namely, the convolutional layer and the sub-sampling or pooling layer. One major advantage of the convolutional neural network is the use of shared filter bank (weights) in the convolutional layers. This improves performance while reducing the size of the parameter set [41]. After the aforesaid layers comes the optional fully connected layer(s) that connects output layer with the complete network. Input to the convolutional neural network is always a 2D or 3D data matrix. The convolutional layer comprises of 'k' number of filters which convolves with the input matrix to produce feature maps. Further, each map is passed on to the sub-sampling layer for max or mean pooling operations. Either before or after pooling operation, each feature map is passed through a non-linear activation function i.e. sigmoid or ReLU etc., to incorporate non-linearity into the feature maps with an output layer follows the convolutional and sub-sampling layers. Moreover, the back-propagation algorithm has been used for weights correction and updation. In this study the optional fully connected layer(s) have not been used as part of the CNN while max pooling operation being used in the sub-sampling layers and sigmoid function being used as the non-linear activation function. Further, spectrograms (Time-Frequency Maps) have been used as inputs to the classifier with time being kept as the receptive field. Figure 2 illustrates the architecture of CNN used in this study.

The convolution operation in the convolutional layer can be mathematically described by the following equation,

$$CFM_j^i = f(\sum_{i=1}^{I} X_i^l \otimes K_{ij}^l + b_j^l) \tag{8}$$

where, 'CFM_j^l' is the output feature map from the l^{th} convolutional layer, 'f' represents the non-linear activation function, 'X_l^l' represents the l^{th} input matrix, 'K_{ij}^l' represents the kernal and 'b_j^l' represents the bias value. Similarly, the pooling operation in the sub-sampling layer can be mathematically described by the following equation,

$$PFM_{i,m}^l = \max_{n-1}^{G} CFM_{i,(m-1)*s+n}^l \tag{9}$$

where, '$PFM_{i,m}^l$' is the output feature map from the l^{th} sub-sampling layer, 'G' represents the pooling size and 's' represents the shifting parameter.

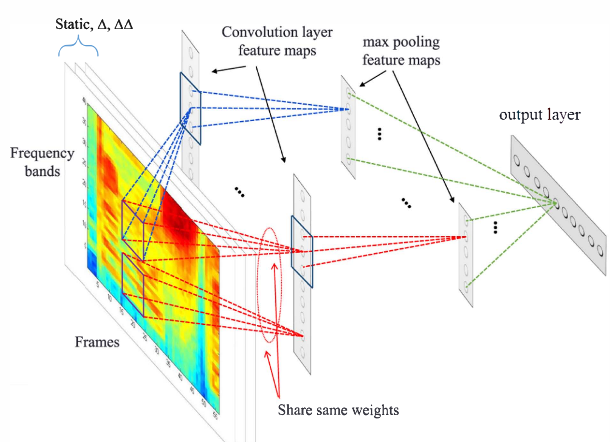

Fig. 2: An illustration of the regular CNN model with convolutional Layer, sub-sampling layer and output Layer

III. Simulation Results and Discussion

All simulations have been conducted in MATLAB. Further, all results have been averaged over 20 iterations with an observed variance of $\pm 0.5\%$. For better and credible analysis of the proposed system, two datasets have been used in this study, that is, a benched mark synthetic dataset downloaded from DOSITS [27], [28] and a raw sonar dataset. In addition, spectral estimates whether 1D or 2D were processed with TPSW method to remove any bias present in the spectral sequence. The system was tested under the effects of noise i.e. additive white Gaussian noise (AWGN) at different levels of signal-to-noise ratio (SNR). Details of simulation parameters along with details of datasets are listed in table I.

TABLE I: Simulation Details

Parameters	Attributes
Sampling Frequency (Hz)	44100
Down-Sampling Factor	10
Sample Size (FrameSize) (sec)	$25ms$
Step-Size (sec)	$12.5ms$
Database	DOSITS [27] and Raw Dataset
Total No. of Samples	DOSITS - 21200 Approx Raw Dataset - 657600 Approx
No. of Classes	DOSITS - 20 Raw Dataset - 04
% Samples for Training	50%
% Samples for Testing	50%
Noise Type	AWGN
SNR Range	0 dB, 10 dB and 10 dB
Feature Learning and Dimension Reduction Techniques	Lofar, PCA, LDA and MPCA
Classifiers	RVLR-MFNN and CNN
Platform	MATLAB

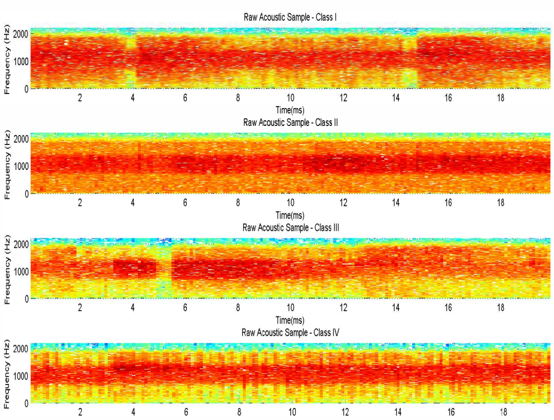

Fig. 3: LOFARogram: waterfall display represents the TF analysis for ships while each waterfall represents one of the 4 respective classes of the raw sonar dataset. In x-axis, time information is shown (from 0 to 25 msec) and in y-axis, frequency information is shown (from DC to 2205 Hz)

Further, dataset collected using passive sonar system has samples belonging to 4 distinct classes of ships. Ship signatures were collected in the Indian ocean. Figure 3 illustrates spectrograms of samples belonging to each one of the four respective classes.

Details of parameters for the two neural networks i.e. RVLR-NN & CNN and MPCA have been listed in table II.

TABLE II: Details of Parameters for RVLR-MFNN, CNN and MPCA

Tech.	Parameters	Values
RVLR MFNN	Activation Functions, ϕ	Tansig & Logsig
	No. of Layers	05
	No. of Hidden Layers	03
	Neurons in Input Layer	Size of FV
	Neurons in Hidden Layers	150-100-100
	Neurons in Out. Layer - DOSITS	20
	Neurons in Out. Layer - Raw Dataset	4
	Learning Rate, η	0.008-0.05
	Alpha, α	0.2
	Gamma, γ	0.00001
	Total Epochs	500
CNN	Activation Function, ϕ	Sigmoid
	Size of the Input (Feature Vector)	98×98
	No. of Hidden Layers	05
	No. of Convolution Layers	02
	No. of Output Maps in Conv. Layers	$20 - 10$
	Size of Kernel	7
	No. of Sub-sampling Layers	02
	Scaling Fac. in Sub-sampling Lay.	02
	Neurons in Output Lay. - DOSITS	20
	Neurons in Output Lay. - Raw Dataset	4
	Learning Rate, η	0.01
	Batch Size	7
	Total Epochs	5000
MPCA	Size of the Input (Feature Vector)	98×98
	No. of Modes	2
	Value of Variation in Each Mode	97
	No. of Iterations (For Optimization)	02
	Size of Prin. Comp. Matrix - Raw Dataset	68×94
	Size of Prin. Comp. Matrix - DOSITS	64×84

Table III and IV gives details of simulation results for both the datasets. Results presented in table III used spectral estimates extracted using LOFAR as feature vector. Moreover, the number of frequency bins were kept to 512 and RVLR-NN was used for classifying objects whereas results presented in table IV used spectrograms (LOFARogram) as well as processed spectrograms as inputs to the classifier. Further, CNN was used to classify objects based on the extracted feature set i.e. LOFARogram and Processed LOFARogram via MPCA. Also, parameter selection for both the classifiers was

made after having rigorous number of simulation runs. Results have been presented using three parameters, that is, accuracy, sensitivity and specificity. Accuracy measures the proportion of correct calls made by the classifier, sensitivity (true positive rate (TPR)) measures the proportion of positives being correctly classified as such whereas specificity (true negative rate (SPC)) measures the proportion of negatives being correctly classified as such by the classifier. Mathematically, the three aforementioned parameters are defined as,
Accuracy (ACC),

$$ACC = \frac{TP+TN}{P+N} = \frac{TP+TN}{(TP+FN)+(FP+TN)}$$

Specificity/ True Negative Rate (SPC),

$$SPC = \frac{TN}{N} = \frac{TN}{FP+TN}$$

Sensitivity/ Recall/ True Positive Rate (TPR),

$$TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$$

TABLE III: %Recognition - LOFAR (Feature Set) - RVLR-NN (Classifier)

Dataset	SNR	% ACC	% TPR	% SPC
DOSITS	0	79.41%	71.40%	76.40%
	10	89.33%	81.23%	87.96%
	20	97.21%	89.29%	94.56%
RSD	0	77.86%	69.68%	73.36%
	10	89.29%	80.48%	86.49%
	20	92.86%	84.86%	89.87%

Figure 4 gives insight to the overall performance of RVLR-NN classifier. Moreover, receiver operating characteristics (ROC) curves for RVLR-NN have been plotted for both the datasets at different noise levels.

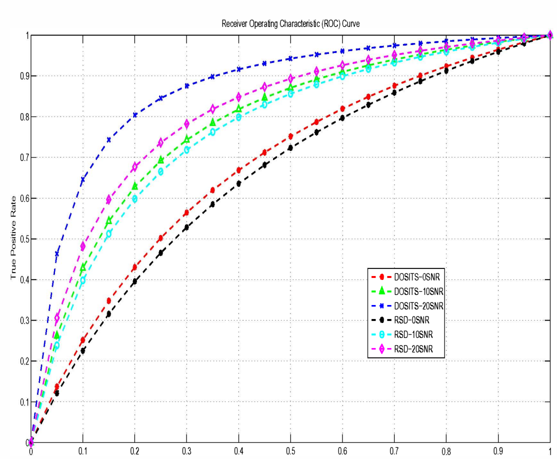

Fig. 4: Receiver Operating Characteristics (ROC) - RVLR-NN

Moreover, area under the curve (AUC) of the ROC curves were also calculated using the composite trapezoidal approach i.e. a numerical method technique for integration. Mathematically, it is defined as,

$$\int_a^b f(x) = (h/2)*(f_a+2f_1+2f_2+\ldots\cdots+2f_{M-2}+2f_M+f_b) \tag{10}$$

where, h is the spacing between two consecutive points.

Area under the curve (AUC) values for the six respective ROC curves in figure 4 are 67.67%, 78.26%, 86.83%, 65.50%, 76.79% and 80.66%, respectively. Further, AUC values suggest that RVLR-NN performs decent classification at low SNR and with the increase in signal-to-noise ratio (SNR) the performance of the classifier also improves.

Results presented in table IV using LOFARogram-CNN have shown drastic improvements in overall classification accuracies compared to the results observed with LOFAR-RVLRNN. Further, the use of MPCA for feature learning has improved the classification rates compared to the results observed with unprocessed LOFARogram-CNN. In addition, with the use MPCA, the size of the feature set (LOFARogram) was also reduced to a minimum of 33% of the original. Table II can be referred for further details about dimensions of the feature set.

TABLE IV: %Recognition - LOFARogram (Feature Set) - CNN (Classifier)

Dataset	Fea. Set	SNR	% ACC	% TPR	% SPC
DOSITS	Spec.	0	95.11%	92.23%	94.12%
		10	97.47%	95.38%	95.90%
		20	98.23%	97.17%	97.98%
	MPCA	0	98.20%	96.38%	97.20%
		10	98.90%	97.20%	98.10%
		20	99.30%	98.11%	98.28%
Raw Sonar Dataset	Spec.	0	96.50%	93.20%	94.23%
		10	98.21%	96.48%	97.58%
		20	98.31%	97.14%	97.95%
	MPCA	0	98.90%	97.28%	98.10%
		10	99.10%	98.58%	98.58%
		20	99.40%	98.90%	98.96%

Figure 5 gives insight to the overall performance of the CNN classifier. Moreover, receiver operating characteristic (ROC) curves for CNN have been plotted for both the datasets with two different feature sets i.e. LOFARogram and processed LOFARogram, as inputs at different levels of signal-to-noise ratio (SNR). Further, area under the curve (AUC) values for the twelve respective ROC curves in figure 5 are 86.47%, 89.28%, 92.95%, 91.47%, 93.18%, 93.57%, 86.70%, 92.16%, 92.89%, 93.18%, 94.19%, and 95.01%, respectively. Further, AUC values suggest that CNN performs exceptionally well with excellent classification rates even at low values of signal-to-noise ratio (SNR). Moreover, combination of MPCA and CNN has outperformed the combination of LOFAR and RVLR-NN in terms of classification accuracies.

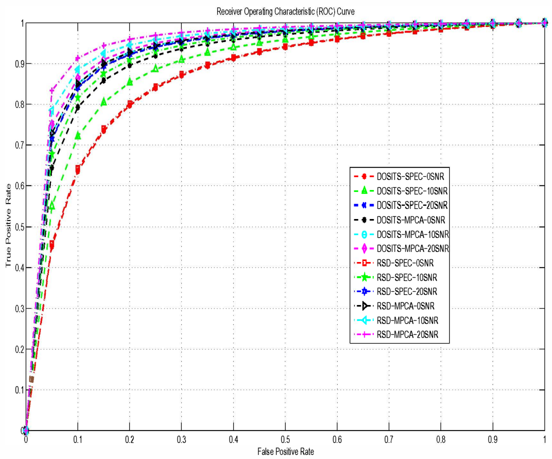

Fig. 5: Receiver Operating Characteristics (ROC) - RVLR-NN

Computational complexity plays an important role in actual realization of any computing system. Moreover, size of the feature vector directly effects the computational cost of the detection and classification system. In this study, effects of dimensionality reduction have been observed on classification rates. For the said purpose, principal component analysis (PCA) and linear discriminant analysis (LDA) have been used for dimensionality reduction of the extracted LOFAR feature set while for classification, RVLR-NN has been used. Further, results have been observed for both the datasets at different levels of signal-to-noise ratio (SNR).

Table V and VI list classification results for dataset downloaded from DOSITS whereas table VII and VIII list classification results for dataset acquired via passive sonar platform. Moreover, original dimension of the feature set was 512 which was varied with a step-size of 25%.

TABLE V: DOSITS - Dimension vs % Recognition - Principal Component Analysis (LOFAR)

		Dimensions - Feature Vector			
		128	256	384	512
SNR	0	53.00%	55.00%	51.40%	57.80%
	10	72.60%	72.60%	72.00%	75.60%
	20	79.20%	87.40%	88.80%	89.40%

Simulation results in table V to VIII depicts that even with a considerable amount of decrease in dimension of the feature vector i.e. to a minimum of 50% of the original, there hasn't been much observed degradation in performance of the classification system. In addition, use of linear subspace learning mechanisms with LOFAR feature set has contributed with an increase in classification accuracies as compared to the results observed while using only LOFAR spectral estimates as feature vector for classification.

TABLE VI: DOSITS - Dimension vs % Recognition - Linear Discriminant Analysis (LOFAR)

		Dimensions - Feature Vector			
		128	256	384	512
SNR	0	55.40%	54.45%	56.80%	59.00%
	10	75.70%	78.00%	79.80%	80.20%
	20	89.70%	91.70%	91.25%	91.65%

TABLE VII: Raw Dataset - Dimension vs % Recognition - Principal Component Analysis (LOFAR)

		Dimensions - Feature Vector			
		128	256	384	512
SNR	0	81.43%	84.43%	83.86%	86.86%
	10	94.14%	84.43%	91.29%	95.00%
	20	92.71%	92.71%	89.14%	85.71%

TABLE VIII: Raw Dataset - Dimension vs % Recognition - Linear Discriminant Analysis (Lofar - Bartlett)

		Dimensions - Feature Vector			
		128	256	384	512
SNR	0	58.21%	67.43%	71.29%	75.75%
	10	84.11%	89.89%	92.93%	94.32%
	20	86.79%	91.07%	98.61%	98.32%

IV. Conclusion

In this study, performance evaluation of linear and multi-linear subspace learning mechanisms have been made for underwater object classification based on underwater acoustics. Further, based on experimental results following conclusions are in order,

1) Time-Frequency metric contains more information about the characteristics of the source and are more robust as compared to be dealing with only spectral estimates (1D) of the signal.
2) Source characteristics are sparsely populated in the spectrum, so it is better to use some transformation with spectral estimates of the source. It may well also reduce the size of the feature vector, which will eventually result in a less complex classification process.
3) Convolutional neural network (CNN) has performed exceptionally well compared to robust variable learning

rate feed-forward neural network (RVLR-NN) in terms of overall operating characteristics of the classifier.

4) Multi-linear principal component analysis (MPCA) has outperformed both the linear subspace learning approaches in terms of overall classification accuracies.
5) Among the two linear sub-space learning techniques, principal component analysis (PCA) has seem to perform better in terms of classification accuracies at low levels of signal-to-noise ratio (SNR).
6) With an increase in signal-to-noise ratio (SNR), considerable increase in system's performance has been observed.

REFERENCES

[1] R. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. Moretti, "Passive acoustic detection and localization of sperm whales (physeter macrocephalus) in the tongue of the ocean," *Applied acoustics*, vol. 67, no. 11, pp. 1091–1105, 2006.

[2] H. Peyvandi, H. Roufarshbaf, M. Farrokhrooz, and S.-J. Park, *SONAR systems and underwater signal processing: classic and modern approaches*. INTECH Open Access Publisher, 2011.

[3] K. W. Chung, A. Sutin, A. Sedunov, and M. Bruno, "Demon acoustic ship signature measurements in an urban harbor," *Advances in Acoustics and Vibration*, vol. 2011, 2011.

[4] T. A. Lampert and S. E. OKeefe, "A survey of spectrogram track detection algorithms," *Applied acoustics*, vol. 71, no. 2, pp. 87–100, 2010.

[5] J. Ghosh, L. Deuser, and S. D. Beck, "A neural network based hybrid system for detection, characterization, and classification of short-duration oceanic signals," *IEEE Journal of Oceanic Engineering*, vol. 17, no. 4, pp. 351–363, 1992.

[6] N. H. Klausner, "Underwater target detection using multiple disparate sonar platforms," 2007.

[7] R. J. Urick, *Principles of underwater sound for engineers*. Tata McGraw-Hill Education, 1967.

[8] C. Kang, X. Zhang, A. Zhang, and H. Lin, "Underwater acoustic targets classification using welch spectrum estimation and neural networks," in *International Symposium on Neural Networks*. Springer, 2004, pp. 930–935.

[9] Y. Zhang, A. B. Baggeroer, and J. G. Bellingham, "Spectral-feature classification of oceanographic processes using an autonomous underwater vehicle," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 726–741, 2001.

[10] N. N. de Moura *et al.*, "Novelty detection in passive sonar systems using support vector machines," in *2015 Latin America Congress on Computational Intelligence (LA-CCI)*. IEEE, 2015, pp. 1–6.

[11] H. Xi-Ying, C. Jin-Fang, H. Guang-Jin, and L. Nan, "Application of bp neural network and higher order spectrum for ship-radiated noise classification," in *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, vol. 1. IEEE, 2010, pp. V1–712.

[12] V. W. Porto, "Evolutionary methods for training neural networks for underwater pattern classification," in *Signals, Systems and Computers, 1990 Conference Record Twenty-Fourth Asilomar Conference on*, vol. 2. IEEE, 1990, p. 1015.

[13] J. De Seixas, N. De Moura *et al.*, "Preprocessing passive sonar signals for neural classification," *IET radar, sonar & navigation*, vol. 5, no. 6, pp. 605–612, 2011.

[14] R. O. Nielsen, *Sonar signal processing*. Artech House, 1991.

[15] U. Demšar, P. Harris, C. Brunsdon, A. S. Fotheringham, and S. McLoone, "Principal component analysis on spatial data: an overview," *Annals of the Association of American Geographers*, vol. 103, no. 1, pp. 106–128, 2013.

[16] G. S. Okimoto and D. W. Lemonds, "Principal component analysis in the wavelet domain: New features for underwater object recognition," in *AeroSense'99*. International Society for Optics and Photonics, 1999, pp. 697–708.

[17] W. Soares-Filho, J. M. De Seixas, and L. Pereira Caloba, "Principal component analysis for classifying passive sonar signals," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, vol. 3. IEEE, 2001, pp. 592–595.

[18] C. Chen, "Recognition of underwater transient patterns," *Pattern recognition*, vol. 18, no. 6, pp. 485–490, 1985.

[19] S. Ji, F. Cong, P. Jia, and Z. Shi, "Principal component analysis on lofar spectrum features of underwater target signals," *J Data Acquis Process*, vol. 18, no. 2, p. 123e6, 2003.

[20] G. Ginolhac and G. Jourdain, """ principal component inverse" algorithm for detection in the presence of reverberation," *IEEE Journal of oceanic engineering*, vol. 27, no. 2, pp. 310–321, 2002.

[21] A. Kundu, G. C. Chen, and C. E. Persons, "Transient sonar signal classification using hidden markov models and neural nets," *IEEE Journal of Oceanic Engineering*, vol. 19, no. 1, pp. 87–99, 1994.

[22] P. W. Van De Ven, T. A. Johansen, A. J. Sørensen, C. Flanagan, and D. Toal, "Neural network augmented identification of underwater vehicle models," *Control Engineering Practice*, vol. 15, no. 6, pp. 715–725, 2007.

[23] Azimi-Sadjadi, M. R., and et al., "Underwater target classification using wavelet packets and neural networks," *Neural Networks, IEEE Transactions on*, vol. 11, no. 3, pp. 784–794, 2000.

[24] C. Chin-Hsing, L. Jiann-Der, and L. Ming-Chi, "Classification of underwater signals using wavelet transforms and neural networks," *Mathematical and computer modelling*, vol. 27, no. 2, pp. 47–60, 1998.

[25] J. E. Simmonds, F. Armstrong, and P. J. Copland, "Species identification using wideband backscatter with neural network and discriminant analysis," *ICES Journal of Marine Science: Journal du Conseil*, vol. 53, no. 2, pp. 189–195, 1996.

[26] T. L. Hemminger and Y.-H. Pao, "Detection and classification of underwater acoustic transients using neural networks," *IEEE Transactions on neural networks*, vol. 5, no. 5, pp. 712–718, 1994.

[27] G. S. o. O. University of Rhode Island, "Discovery of sound in the sea." [Online]. Available: http://www.dosits.org

[28] K. A. Cunningham and D. C. Mountain, "Simulated masking of right whale sounds by shipping noise: Incorporating a model of the auditory periphery," *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1632–1640, 2014.

[29] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Mpca: Multilinear principal component analysis of tensor objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, 2008.

[30] T. Aboulnasr and K. Mayyas, "A robust variable step-size lms-type algorithm: analysis and simulations," *Signal Processing, IEEE Transactions on*, vol. 45, no. 3, pp. 631–639, Mar 1997.

[31] M. Baqar, S. Azhar, Z. Iqbal, I. Shakeel, L. Ahmed, and M. Moinuddin, "Efficient iris recognition system based on dual boundary detection using robust variable learning rate multilayer feed forward neural network," in *Information Assurance and Security (IAS), 2011 7th International Conference on*. IEEE, 2011, pp. 326–330.

[32] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[33] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[34] L. Deng, J. Li, J.-T. Huang, K. Yao, D. Yu, F. Seide, M. Seltzer, G. Zweig, X. He, J. Williams *et al.*, "Recent advances in deep learning for speech research at microsoft," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2013, pp. 8604–8608.

[35] G. M. Wenz, "Acoustic ambient noise in the ocean: spectra and sources," *The Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1936–1956, 1962.

[36] J. Di Martino, J.-P. Haton, and A. Laporte, "Lofargram line tracking by multistage decision process," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on*, vol. 1. IEEE, 1993, pp. 317–320.

[37] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. AU-15, p. 7073, 1967.

[38] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[39] B. Scholkopft and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX*, vol. 1, no. 1, p. 1, 1999.

[40] J.-T. Huang, J. Li, and Y. Gong, "An analysis of convolutional neural networks for speech recognition," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 4989–4993.

[41] O. Abdel-Hamid, A.-R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, "Convolutional neural networks for speech recognition," *IEEE/ACM Transactions on audio, speech, and language processing*, vol. 22, no. 10, pp. 1533–1545, 2014.