Defence Science Journal, Vol. 71, No. 1, January 2021, pp. 117-123, DOI : 10.14429/dsj.71.14929

Passive Shallow Water Automated Target Recognition using Deep Convolutional Bi-directional Long Short Term Memory

Suraj Kamal*, C. Satheesh Chandran, and M.H. Supriya

Department of Electronics, Cochin University of Science and Technology, Kochi - 682 022, India
**E-mail: surajkamal@cusat.ac.in*

ABSTRACT

The extremely challenging nature of passive acoustic surveillance makes it a key area of research in Naval Non-Co-operative Target Recognition especially in Anti-Submarine Warfare systems. In shallow waters, the complex acoustics due to the highly varying ambient background noise as well as the multi-modal propagation in the surface-bottom bounded channel makes surveillance even difficult. In this work, an ensemble of Convolutional Neural Networks and Bidirectional Long Short Term Memory stages employing soft attention is used to effectively capture the spectro-temporal dynamics of the target signature. In order to alleviate the overall computational cost associated with the optimal model search in the extensive hyperparameter space, a recursive model elimination scheme, making frugal use of the available resources, is also proposed. Experimental analysis on acoustic target records, collected from the shallows of Arabian Sea, has yielded encouraging results in terms of model accuracy, precision and recall.

Keywords: Shallow waters; Passive sonar; Automated target recognition; CNN; LSTM

1. INTRODUCTION

With the emergence of many asymmetric forces, stealthier platforms, proliferation of unmanned surface and submerged platforms, the naval conflict zones have mostly shifted from deep ocean towards the littoral zones during the past two decades[1]. However, in shallow waters, reliable target recognition is formidably hard, manifested by the rather complex acoustics, multi-modal propagation in the surface-bottom bounded channel, complex hydro-meteorological conditions, highly varying ambient background noise as well as comparatively low source levels[2,3]. Various spatio-temporal inhomogeneities in the shallow water medium makes it a dispersive stochastic filter, which can often mask the signals of interest beyond the detection threshold[4]. The cacophony of marine biologics produced by large schools of marine life such as snapping shrimps in the littorals of tropic and subtropic waters could easily overpower even a medium scale frigate's radiated noise in near fields.

Detection of envelope modulation on noise (DEMON) as well as low frequency analysis and recording (LOFAR) are canonical examples of spectral features exploited in open ocean target detection and classification. However, the prevalence of non-target contacts and the mutual interference of broadband components due to surface-bottom interactions can jeopardize the success rates of classifiers based on such features, while operating in shallow waters. Das[5], *et al.* proposed a cepstral features based classifier to reduce these distortion effects. Kuperman[6], *et al.* suggests developing methods that make use of the data themselves in order to alleviate these challenges while confronting shallow water acoustics.

Recent developments in artificial intelligence (AI) and machine learning (ML) have enabled end-to-end learning, often referred to as deep learning[7], a special variant of artificial neural networks (ANN). While classical multi-layer perceptrons (MLPs) have a single hidden representation layer, deep neural networks (DNNs) create multiple levels of hierarchical abstraction within the network itself, yielding better invariant representations at the higher layers. DNNs often eliminate the requirement of intrinsic hand engineered features[8]. Instead, the network learns from the raw data or from an intermediate representation that well preserves the latent structures in the data.

In this paper, a machine learning based naval non-co-operative automated target recognition (NCATR) system is proposed, which can effectively mitigate several adversarial effects presented by the shallow water acoustics in performing target recognition. The system utilises a deep convolutional neural network[8] (CNN) for supervised feature learning, in conjunction with a bidirectional long short term memory[9] (BLSTM) employing soft attention[10] mechanism. A recursive model elimination strategy is also proposed, which can effectively minimise the model space introduced by the excessive number of hyperparameters involved in the network design.

Received : 08 September 2020, Revised : 21 December 2020
Accepted : 23 December 2020, Online published : 01 February 2021

2. SYSTEM IMPLEMENTATION

The passive sonar listens to the radiated target signatures in its vicinity using an array of hydrophones. The raw acoustic signal captured by the wet end of the sonar is transformed into a time-frequency (TF) representation such as spectrogram in order to make the spectral as well as temporal dynamics of the signal more explicit. The salient spectral region of targets of interest spans low frequencies ranging from a few hertz to 5 kHz approximately[11]. Hence, in addition to using raw spectrogram (spgm) as the basis representation, a log-scaled spectrogram (logspgm) is also employed in order to make the low frequency spectrum more dilated and to expose the hidden subtle tonals. This is quite evident in Fig. 1, which depicts the spectrogram and log-scaled version of the same, corresponding to an observed target signature.

2.1 The Proposed Network Architecture

The architecture of the network is as depicted in Fig. 2. CNNs are a specific variant of the ANNs, introducing the concept of receptive fields, weight reuse and local pooling of features. Hence, CNNs are used as the spectro-temporal feature learners at the initial layers in order to ensure invariance both in time and frequency. A multi-filter convolution approach is used at the input layer, with varying temporal and spectral dimensions in order to better capture the different temporal and spectral dynamics of the target signature. A non-linear activation function given by *max* (0, *a*), often termed as rectified linear unit[12] (ReLU), can be applied at the CNN pre-activation *a* to yield the activation map. In the current work, either of the two variants of ReLU, known as the Leaky ReLU (LReLU) and the Parametric ReLU (PReLU), which offers a better flow of gradients, is utilised as the activation function.

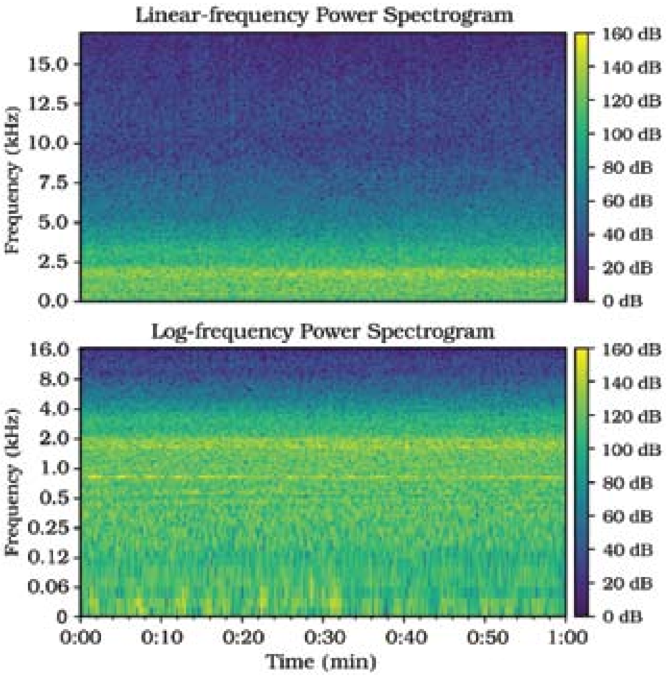

Figure 1. Spectrogram and log-spectrogram corresponding to a target's acoustic signature.

A non-parametric operation known as pooling[13] is often used in conjunction with the convolutional layer to reduce the number of parameters as well as to improve the invariance by estimating either the maximum or average under local patches of the activation map. In this work, a symmetric max pooling is used across all pooling layers. During the gradient descent process, the internal weight distribution can be severely altered, leading to internal covariate shift. A batch normalisation[14] layer,

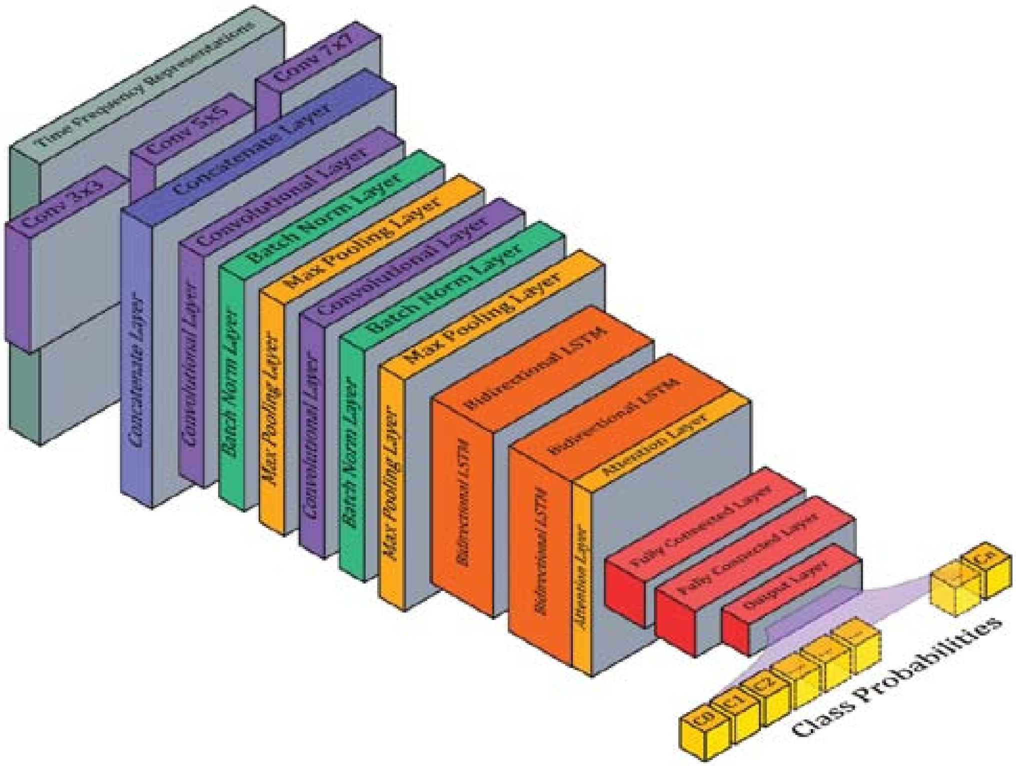

Figure 2. The proposed network architecture.

which normalises the weights by their mean and covariance, is employed throughout the network to reduce such adversarial effects.

The spectro-temporal features learned by the CNN alone might not be able to encode the fine sequential temporal dependencies present in the observed target signature. The extreme variabilities and several adversarial phenomena in shallow waters demand the exploitation of latent temporal cues also, in order to improve the classifier performance. Recurrent Neural Network (RNN) provides a way to learn sequential dependencies in the signal by virtue of its structure. Long Short Term Memory (LSTM) is a special kind of RNN that ameliorates the vanishing gradient problem over long temporal steps. However, a conventional unidirectional LSTM cell makes prediction only in the context of past spectro-temporal representations. Since the acoustic signature of an underwater target has quite complex sequential dependencies, it will be advantageous to compute backwards also so as to make the latent sequential relationships in the data more explicit. This can be achieved with the help of bidirectional LSTMs, having forward and backward recurrent units in two separate hidden layers, for the forward as well as backward computation respectively.

Since the acoustic signature of an underwater target has quite complex sequential dependencies, a BLSTM with peephole architecture[15] is employed as the temporal sequence learner. In order to reduce the effects of spurious and sporadic variations as well as to focus on significant spectro-temporal features, a soft attention module is also incorporated in the BLSTM stage. The attention function is implemented as a separate importance weighted network, which is differentiable against the cost function, so that it could be easily trained with Stochastic Gradient Descent (SGD). In the fully connected output layer, a softmax activation is used in order to express the network output as a multi-nomial distribution over all known classes.

The hierarchical spectro-temporal feature learning is performed in a supervised manner using the deep CNN with error back propagation. Cross entropy[16] or the negative log-likelihood between the prediction and the true target is employed as the cost function. The classifier is optimised using a modern variant of SGD, known as the Adaptive Momentum[17] (ADAM) optimiser, since it provides better numerical stability by adoptively regulating the learning rate during the gradient descent process.

3. EXPERIMENTAL SETUP

3.1 Dataset

The field data required for the training and evaluation of the classifier system have been collected from the shallows of Arabian Sea, the north-western part of Indian Ocean, in the vicinity of international shipping lane off Cochin. Through several expeditions of the research vessel Sagar Sampada and using various hired platforms such as survey boats, multiple custom-built buoys together with horizontal line arrays have been deployed for carrying out the acoustic recordings. The raw observations have been made at 96 ksps and 24 bps using up to five omni-directional hydrophones having input sensitivity of -180 dB re 1 V/µPa with lateral as well as vertical separation of few meters in order to introduce limited but beneficial spatial variance. Recordings have been taken in different sea states and at various locations with diverse depth profiles ranging from 6 m to 50 m and the proximity of the transshipment port has helped in gathering many target records in different ambient noise conditions as well. The constant crackling and other biologics present in the shallows have been removed manually during pre-processing. The acoustic records have been normalised and down sampled to 44.1 ksps as no significant features were observed in the spectrogram above 20 kHz.

Most of these recordings have been made at the point of transit of the target platforms along the shipping lane at various cruising speeds ranging from 2 kn to 15 kn with the Closest Point of Approach (CPA) of about a few hundred meters. From the repertoire of the collected acoustic records, a set of 31 targets belonging to 10 broad categories including biologics, has been chosen for evaluating the classifier. The acoustic records have been partitioned into three disjoint sub sets, the training set, the validation set and the test set with randomly selected partitions from the time domain having 60%, 20% and 20% of the total duration of the records correspondingly. Altogether the target records have an on-disk size of 31 GB approximately.

3.2 Recursive Model Elimination

Apart from the parameters trainable through backpropagation, a neural network has a multitude of non-trainable parameters, often termed as hyperparameters such as the number of layers, LSTM nodes, kernel size, regularisation factor, learning rate, activation function and optimisation algorithm to name but a few. The permutations and combinations of these parameters in effect are staggeringly high, which makes it a combinatorial search problem and is often expensive in terms of computational resources. Hence a recursive model elimination scheme is proposed here for effectively reducing the number of model configurations. The parameters chosen for optimisation are listed in Table 1, where nCNN represents the number of convolutional feature learning layers, nLSTM denotes the number of LSTM layers, Reg indicates whether regularisation is used, Attn indicates whether a soft attention layer is employed, BS is the batch size, TF is the time-frequency representation used, LR is the learning rate and Act is the type of activation applied.

Considering $\rho_1, \rho_2, ..., \rho_n$ as the hyperparameters to be optimised and P_1, P_2, … , P_n as their respective domains, the corresponding number of configurations in the hyperparameter space can be defined as,

Table 1. Hyperparameters for optimisation

nCNN	nLSTM	Reg	Attn	BS	TF	LR	Act
2	1	True	True	256	spgm	1e-3	relu
3	2	False	False	512	logspgm	1e-4	lrelu
					both	1e-5	prelu
$P_{13} = 2$	$P_{14} = 2$	$P_{15} = 2$	$P_{16} = 2$	$P_{17} = 2$	$P_{18} = 3$	$P_{19} = 3$	$P_{20} = 3$
							P = **864**

$$P = \prod_{i=1}^{n} P_i \tag{1}$$

so that the model space $M=\{m_1, m_2, \ldots, m_P\}$. In the current scenario, $P = 864$. The optimisation problem for a model m_i on a set of examples S, which can be either training or validation instances sampled from a latent probability distribution D_x, can be expressed as,

$$\hat{\rho} = \arg\min_{\rho\in P} \mathrm{E}_{x\sim D_x}[L(x; m_{i\rho}(S))] \tag{2}$$

The optimisation is carried out recursively in multiple passes to minimise the expected loss L. The maximum validation accuracy for each model in the entire training epochs, evaluated for all the models after each pass j, can be expressed as,

$$\eta_j = acc_j \forall_m \left\{ \max_{i\in e}(m(i)_{s^{valid}}) \right\}; j = 1\ldots n \tag{3}$$

where acc denotes the accuracy and n is the number of passes. Models with $\eta_j < \eta_T$ are eliminated after each pass, where η_T = 90% and 94% respectively for $j = 1, 2$. The model parameters for the top 10 models obtained after pass 1 of 50 epochs, along with their corresponding accuracies for pass 2 and 3, sorted based on validation accuracy, are shown in Table 2. It is observed that the models which do not employ regularisation have a clear tendency to overfit after 150 epochs. Hence these models are also eliminated in pass 3 even though they have sufficiently high accuracies. The entire process of model training and elimination has been performed on an NVIDIA GPU cluster with an aggregate computational capability of 80 TFLOPS approximately. Python together with CUDA[18] has been used for developing the classifier models.

4. RESULTS AND DISCUSSIONS

After all the elimination passes, three models, viz., Model 3, Model 2 and Model 4 have been identified as the candidates for deployment. Although the margin of improvement obtained by the soft attention module isn't quite noticeable on accuracy, it has consistently performed better throughout the entire model elimination passes and has hence remained in the list of final candidates. It is still interesting to note that for Model 3 and Model 4 having the soft attention module, the classification accuracy for the worst performing class i.e. class 23, is considerably better than the classifier without attention. This can be clearly observed in the per class accuracy plot depicted in Fig. 3. The training, validation and test accuracies along with their corresponding losses for these three models are plotted in Fig. 4. The test accuracies of the respective models are observed to be 94.29%, 94.42% and 94.32%.

Table 2. Models selected after elimination pass 2 and 3

Model	Input Parameters			Network Parameters			Accuracies (%)			
	TF		Dim	nCNN	Reg	Attn	Pass 2: Epochs=150		Pass 3: Epochs=250	
	spgm	logspgm					Train	Val	Train	Val
1	✓	✓	64 x 64	3	×	×	99.28	94.96		
2	✓	✓	64 x 128	2	✓	×	93.24	94.95	95.95	94.95
3	✓	✓	64 x 128	2	✓	✓	93.24	94.93	93.95	95.21
4	✓	✓	64 x 64	3	✓	✓	93.84	94.92	94.48	94.84
5	✓	✓	64 x 128	2	×	✓	99.33	94.91		
6	✓	✓	64 x 128	2	×	×	99.32	94.91		
7	✓	✓	64 x 64	2	×	✓	99.34	94.89		
8	×	✓	64 x 128	2	×	✓	99.37	94.83		
9	✓	✓	64 x 128	3	×	✓	99.25	94.79		
10	✓	✓	64 x 64	2	×	×	99.34	94.75		

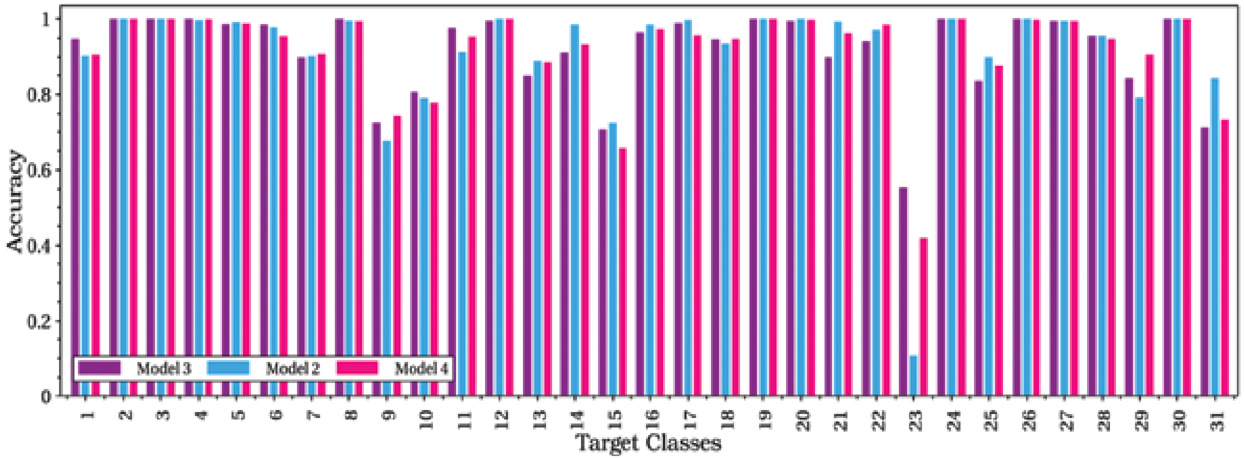

Figure 3. Normalised per class test accuracy for the 31 member classes of the final candidate classifiers.

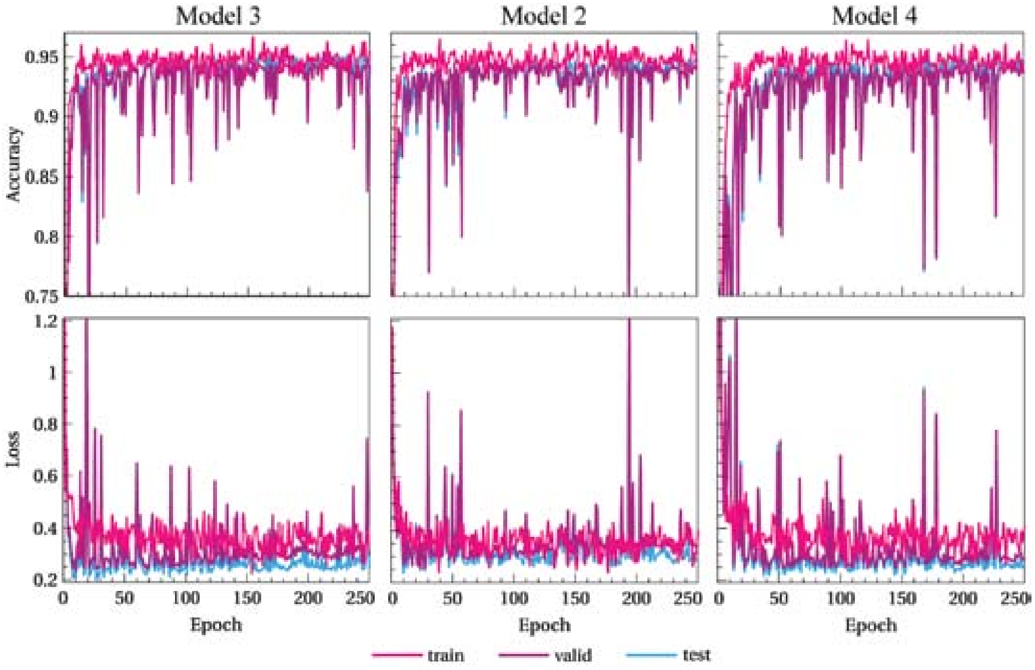

Figure 4. Performance summary of the best candidate models after pass 3.

It is quite evident from Fig. 3 that most of the targets achieve a success rate of over 90%, while certain classes fall short, class 23 for instance. This may be due to the proximity of the embeddings that persist in the feature space learned by the network, which leads to perplexities in the decision space. This can be observed in Fig. 5, where the entangled classes are confused with each other, with class 23 being mostly misclassified as class 7.

Despite the fact that the models have good performance in terms of accuracy, it can be often misleading in the case of a target classifier because a model that classifies all targets as friendly/neutral generally will have an exceedingly high accuracy even if the few true contacts are misclassified. Especially in the territorial waters, where the overwhelming majority of acoustic sources are either friendly or neutral, and the adversarial targets being sparse, the accuracy alone cannot be a reliable metric for assessing the model performance. False alarms can be triggered by two scenarios, one in which a true contact may be misclassified as neutral and the other in which a neutral target may be misclassified as adversarial. Hence, the classifier performance has also been analysed in the light of other metrics such as precision and recall that can take into account the situations of such false alarms.

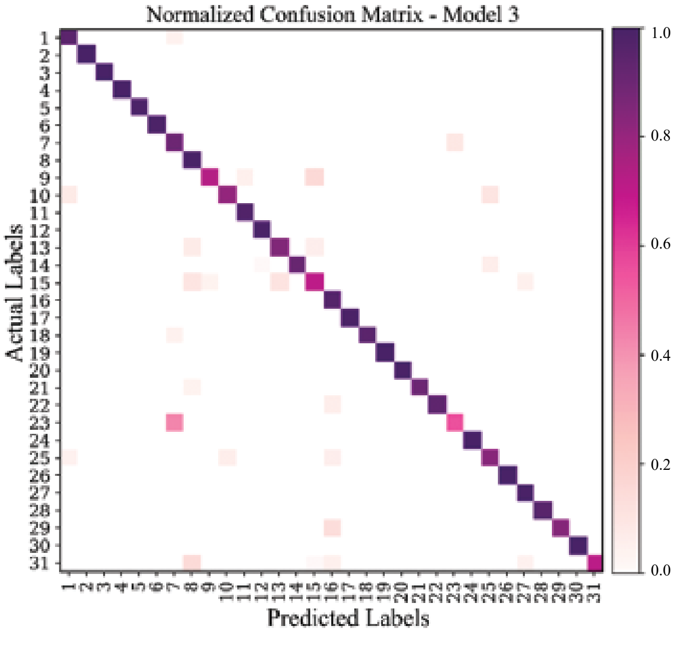

Figure 5. Normalised confusion matrix of Model 3 on test set.

Recall expresses the ability of the classifier to find the few but all the potentially hostile targets from a set of acoustic sources, whereas precision refers to the ratio of targets recognised as hostile that were actually hostile**.** In the context of a shallow water target classifier, recall may be slightly more important, as even if the predicted class is a false positive, it can be confirmed on further investigations using other modalities if possible. A false negative, on the other hand, does not give a chance for this detailed analysis. Precision-Recall Curve (PRC) is an often-recommended metric while classifying severely unbalanced observations and in situations where true negatives i.e. the neutral targets, are not much of a concern.

The PRC for the candidate classifier Model 3 is plotted in Fig. 6, from which it can be observed that the Area Under the Curve (AUC) is comparatively low for class 15 and up to a certain extent for class 7, 9, 13 and 31. It is found that class 23 is the worst performing class, which is in quite agreement with the results presented in Fig. 3 and Fig. 5. The plot also indicates that the F1 scores for the majority of classes are concentrated around the area spanned by F1 = 0.8, which suggests that there is an appreciable balance between precision and recall. Hence the probability of the classifier to miss a hostile target is extremely low, which makes it a justifiable choice.

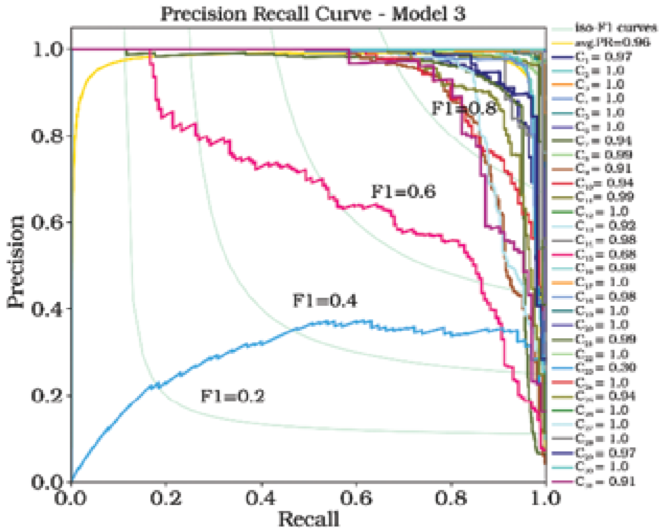

Figure 6. Precision vs. recall for Model 3.

5. CONCLUSIONS

In this work, CNN based feature learning layers followed by temporal sequence learning BLSTMs with a differentiable soft attention module have been employed for NCATR task in shallow waters. All the models have been trained with the archived acoustic target signatures collected from the littorals. In order to reduce the exhaustive model space rendered by large number of hyperparameters, a resource-aware recursive model elimination scheme has been employed. Along with accuracy, measures such as precision, recall and F1 score have also been utilised to evaluate the final candidate models' performance. From the PRC, it has been found that almost all the classes have good AUC with an average PR value of 0.96, which is quite promising. This further confirms the reliability of the proposed classifier model.

REFERENCES

1. Milan, V. The Right Submarine for Lurking in the Littorals. *Proc. - U. S. Nav. Inst.,* 2010, **136**(6), 16-21. http://usni.org/magazines/proceedings/2010/june/right-submarine-lurking-littorals (Accessed on 05 October 2018).
2. Katsnelson, B.; Petnikov, V. & Lynch, J. Fundamentals of shallow water acoustics. Springer Science & Business Media, 2012, pp. 540.
3. Das, Arnab. Shallow ambient noise variability due to distant shipping noise and tide. *Applied acoustics.,* 2011, **72**(9), 660-664.
   doi: 10.1016/j.apacoust.2011.03.003
4. Suganthbalaji, R.; Elizabeth, N.X.; Nair, N. & Nair, P. Effect of Environment on Underwater Acoustic Communication Data Rates. *Def. Sci. J.,* 2019, **69**(2), 163-166.
   doi: 10.14429/dsj.69.14227
5. Das, A.; Kumar, A. & Bahl, R. Marine vessel classification based on passive sonar data: the cepstrum-based approach. *IET Radar, Sonar Navig.,* 2013, **7**(1), 87-93.
   doi: 10.1049/iet-rsn.2011.0142
6. Kuperman, W.A. & Lynch, J.F. Shallow-water acoustics. *Physics Today,* 2004, **57**(10), 55-61.
   doi: 10.1063/1.1825269
7. Hinton, G.E.; Osindero, S. & Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Computation,* 2006, **18**(7), 1527-1554.
   doi: 10.1162/neco.2006.18.7.1527
8. LeCun, Y.; Bengio, Y. & Hinton, G. Deep learning. *nature,* 2015, **521**(7553), 436-444.
   doi: 10.1038/nature14539
9. Graves, A. & Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks,* 2005, **18**(5-6), 602-610.
   doi: 10.1016/j.neunet.2005.06.042
10. Bahdanau, D.; Cho, K. & Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv,* 2014. arXiv:1409.0473.
11. Urick, R.J. Principles of underwater sound. McGraw-Hill, New York, 1975, pp. 384.
12. Nair, V. & Hinton, G.E. Rectified linear units improve restricted boltzmann machines. *In* Proceedings of the 27th international conference on machine learning, 2010, pp. 807-814.
13. Sai, M.; Upadhyay, P. & Srinivasan, B. Fault detection and isolation in electrical machines using deep neural networks. *Def. Sci. J.,* 2019, **69**(3), 249-253.
    doi: 10.14429/dsj.69.14413
14. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *In* Proceedings of the International conference on machine learning, 2015, pp. 448-456.
15. Gers, F.A.; Schraudolph, N.N. & Schmidhuber, J. Learning precise timing with LSTM recurrent networks. *J. Mach. Learn. Res.,* 2002, **3**(Aug), 115–143.
    doi: 10.1162/153244303768966139
16. Rubinstein, R.Y. & Kroese, D.P. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation and machine learning. Springer Science & Business Media, 2004, pp. 301.
17. Kingma, Diederik P. and Jimmy Ba. Adam: A method for stochastic optimization.
    arXiv, 2014. arXiv:1412.6980.
18. Cook, S. CUDA Programming: A Developer's Guide to Parallel Computing with GPUs. Morgan Kaufmann Publishers Inc., San Francisco, USA, 2012, pp. 608.

CONTRIBUTORS

Mr Suraj Kamal completed his Master of Science Programme in Electronics from College of Applied Science, Thodupuzha, affiliated to MG University, Kerala, India, in 2006. He is presently pursuing his PhD in the Department of Electronics, Cochin University of Science and Technology, Kerala. His main areas of research are Applied Underwater Acoustics, Artificial Intelligence and Machine Learning.
He has conceptualised and implemented the models as well as the training system used in this work.

Mr Satheesh Chandran C. completed his Master of Technology in Electronics from Department of Electronics, Cochin University of Science and Technology, Kerala, India, in 2013 and is currently pursuing his Ph.D. in the same department. His current areas of research include Underwater Acoustics & Signal Processing, Machine Learning and Pattern Recognition.

In the present work, he has helped in post-processing and archiving of the acoustic data as well as in creating time-frequency representations for the classifier model.

Dr Supriya M.H. obtained her PhD from Department of Electronics, Cochin University of Science and Technology, Kerala, India, in 2008 and is currently working as Professor in the same department. She has more than 22 years of teaching experience and 4 years of Industrial experience. Her research and teaching areas of interest include Sonar Technology, Signal Processing as well as Underwater Target Recognition.

She has guided the current work as well as helped in obtaining the financial aid for field data collection.