

Deep Learning-Based Modulation Classification Using Time and Stockwell Domain Channeling

Shrishail M Hiremath, Sambit Behura
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela
Rourkela, India
Email: hiremaths@nitrkl.ac.in,
sambitbehura001@gmail.com

Subham Kedia
Department of Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, India
Email: subhamkedia799@gmail.com

Siddharth Deshmukh
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela
Rourkela, India
Email: deshmukhs@nitrkl.ac.in

Sarat Kumar Patra
Indian Institute of Information Technology Vadodara
Vadodara, India
Email: skpatra@iiitvadodara.ac.in

Abstract— Deep learning techniques have recently exhibited unprecedented success in classification problems with ill-defined mathematical models. In this paper, we apply deep learning for RF data analysis and classification. We present a novel method of using I-Q time samples to form images with ‘Time and Discrete Orthonormal Stockwell Transform Domain Channels’ which are used for training a convolutional neural network (CNN) for radio modulation classification. Also, a concept inspired from transfer learning is used in extending the number of output classes of the CNN, which helps the network to estimate the approximate SNR of the input signal as well and further improve the classification accuracy. Such a network trained on Time and Stockwell Channeled Images performs superior to similar networks that are trained on just raw I-Q time series samples or time-frequency images, especially when training samples are less. The network achieved an overall classification accuracy of 97.3% at 8 dB SNR over a class of 10 radio modulation schemes (for both digital and analog systems). The study shows that such a trained network can be well applied to achieve high classification accuracies at low and moderate SNR scenarios.

Keywords—deep learning, convolutional neural network, modulation classification, Stockwell transform, discrete orthogonal Stockwell transform, time and Stockwell domain channeling.

I. INTRODUCTION

RADIO data available from an antenna is often easily captured, but in the modern day it is difficult to label and curate the data accurately from the complex high-data rate RF information. The strategies adopted for such tasks are often time-consuming, and their implementations are not precise under varying environmental conditions. Hence, blind radio signal recognition and identification at the receiver end has turned out to be a very useful and important tool in dense, multi-user scenarios. Fast labeling and understanding of the radio spectrum can provide added advantages like optimized spectrum utilization, minimized and identifiable interference, spectrum policy enforcement, and implementing efficient spectrum sensing and coordination systems. Hence it has enabled radio fault detection, spectrum interference monitoring, dynamic spectrum access, opportunistic mesh networking and numerous other fields in communication systems.

Modulation classification is the process of blindly identifying and differentiating radio signals at the receiver end, as a step towards understanding the type of communication schemes being used by the transmitters in the vicinity. Modulation recognition or classification is a front-end tool in number of applications like link adaptation, modern military signal intelligence systems, spectrum monitoring systems, unmanned aerial drones, dynamic spectrum access, cognitive radio, and cellular standards like LTE-Advanced.

The last two decades have seen wide variety research on developing novel methods and algorithms for automatic radio modulation classification/recognition. Many of these are carefully designed feature extraction based techniques which project the received signal on a low-dimensional feature space in which compact decision boundaries can help differentiate one radio modulation from the other [1]. Modulation classification can also be performed using generative algorithms based on probabilistic models like Naïve Bayes [2], hidden Markov models obtained with maximum likelihood estimation [3] and methods that uses likelihood ratio tests such as the average likelihood ratio test (ALRT) [4], generalized likelihood ratio test (GLRT) [5] and the hybrid likelihood ratio test (HLRT) [6]. Integrated cyclic moment-based features [7] and features based on CSS (Concatenated Sorted Symbols) [8] are widely popular for forming analytically derived decision trees to sort modulations into different classes. Despite its robustness against noise and interference, cyclostationary analysis of a signal has a high computational cost and is not efficient for quick labeling and real-time modulation classification [7].

In the past few years, there have been massive improvements and developments in neural network architectures and optimization algorithms. Deep neural networks have pushed performance boundaries of machine learning tasks in a variety of applications. This deep learning trend, which is quite popular in computer vision or text processing, is yet to be adequately explored and fully applied to complex temporal radio signal datasets. Moreover, in case of RF data, the type of samples that the network is to be trained on, and whether or not some kind of pre-processing on the time samples might improve

training performance needs to be investigated as well. In this work we propose a novel method of using I-Q time samples to form images with ‘Time and Discrete Orthonormal Stockwell Transform Domain Channels’ which are then used for training a convolutional neural network (CNN) for the task of radio modulation classification. This type of training on the network proves to be really efficient, especially when the number of training samples is less.

The organization of this paper is as follow: Section II mentions some notable papers that have used deep neural networks for modulation classification, and also gives a brief background on the need for time-frequency analysis and the Stockwell transform. Section III presents an overview of the available radio machine learning datasets. Section IV presents the details about the proposed classification approach i.e. system model, Time & Stockwell Domain Channeling, the CNN architecture and the concept of Extended Output Classes. Section V presents the experimental results and their detailed analysis. Section VI concludes this paper.

II. BACKGROUND

A. Deep Architectures for RF Data Classification

Applying deep learning to a problem like modulation classification involves selecting a network architecture and hyper-parameters, training the network to optimize weights that minimize loss, and applying the trained network to the problem at hand. [9] presents a survey of the various deep learning architectures inspired from computer vision and natural language processing that can be applied to the task of modulation classification. Some deep architectures that have found success in radio signal identification include Convolutional and Residual Networks [10] [11], Recurrent and LSTM networks [12] and Heterogeneous Deep Fusion Models [13].

B. Need for Time-Frequency Analysis – The Stockwell Transform

One major drawback of Fourier transform (FT) as a spectral analysis tool is that it produces only the time-averaged spectrum, thus making it unfavorable for applications where local information is preferred (e.g., signal de-noising, compression, phase analysis) [14]. Thus in recent years, more advanced representations known as joint time-frequency representations have been adopted [15].

The wavelet transform [16] is a time-frequency decomposition that applies local decomposition filters to a signal on multiple scales. But, even though the term “scale” can be approximately interpreted as “frequency,” there is no way to extract proper frequency information from the scale information [14].

The Stockwell transform (ST, also popularly known as the S-transform) [17] [18] [19] [15] is a time-frequency decomposition that provides absolutely referenced phase information. Here, the summation of the coefficients for a fixed frequency gives the exact Fourier coefficient for that frequency [14].

Consider a one dimensional signal $x(t)$. The ST of $x(t)$ is defined as the FT of the product between $x(t)$ and a Gaussian window function.

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-j2\pi f t} dt \quad (1)$$

By properties of the Gaussian function, the relationship between $S(\tau, f)$ and $X(f)$ (FT of $x(t)$) is given as

$$\int_{-\infty}^{\infty} S(\tau, f) d\tau = X(f) \quad (2)$$

Hence, the summation of Stockwell coefficients along the time axis gives the FT of the signal. The original signal $x(t)$ can be recovered by calculating the inverse FT of $X(f)$ [14].

However, it is known that the Stockwell transform (also discrete ST) is highly redundant and thus it needs a large amount of time and storage space even for a moderately long signal. For example, a signal of length N , generates N^2 coefficients through the discrete ST. As a solution to reduce this redundancy, the time-frequency domain can be partitioned into N regions, and each region can be represented by one coefficient [14]. This is the strategy adopted by the discrete orthonormal Stockwell transform (DOST) [20], thus making its computation simpler. The DOST coefficients can be computed by taking the vector dot-product of the input signal with a set of N basis vectors, which gives it a computational complexity $O(N^2)$ [14]. Let a region in the time-frequency domain be described by a set of parameters: v specifies the center of each frequency band, β is the width of the band and τ specifies the point in time. Using these parameters the k^{th} basis vector is defined as

$$D[k]_{[v,\beta,\tau]} = \frac{1}{\sqrt{\beta}} \sum_{f=v-\frac{\beta}{2}}^{v+\frac{\beta}{2}-1} e^{-j2\pi\frac{k}{N}f} e^{j2\pi\frac{\tau}{\beta}f} e^{-j\pi} \quad (3)$$

For $k = 0, \dots, N-1$. An algorithm to compute DOST through a fast method is presented in [14].

III. RADIO MACHINE LEARNING DATASETS

All datasets used in this work are provided by DeepSig Inc., and are licensed under the Creative Commons Attribution – NonCommercial – ShareALike 4.0 License (CC BY-NC-SA 4.0). DeepSig has created some standard datasets which can be used by scientists and engineers for original and reproducible research. These datasets give scope to machine learning researchers to dive directly into new and important technical areas in radio signal processing without the need for collecting or generating new datasets [21].

Dataset RadioML 2018.01A by DeepSig Inc. includes both synthetic simulated channel effects and over-the-air recordings of 24 digital and analog modulation types which has been heavily validated. This dataset was used in [11] which provides additional details and description of the dataset. Data are stored in hdf5 format as complex floating point values, with 2 million examples, each 1024 samples long. The included modulation classes are 32PSK, 16APSK, 32QAM, FM, GMSK, 32APSK, OQPSK, 8ASK, BPSK, 8PSK, AM-SSB-SC, 4ASK, 16PSK, 64APSK, 128QAM, 128APSK, AM-DSB-SC, AM-SSB-WC, 64QAM, QPSK, 256QAM, AM-DSB-WC, OOK and 16QAM [21].

IV. PROPOSED CLASSIFICATION APPROACH

A. System Model

Fig. 1 presents the complete system model that is being used for our classification approach. A RF I-Q image is a matrix containing fixed number of samples (1024 in this case) of the in-phase and quadrature-phase components of the received signal, arranged into rows. This 2×1024 matrix/image serves as the input to the system model. The DOST block performs row-wise DOST (separately for I-samples and Q-samples) on its input image and then takes the absolute values of each transformed complex values. The output of this block is another 2×1024 image which contains Time-Frequency Domain information about the I and Q samples. The input and output images of the DOST block are then fed into the Time and DOST Data Channeling block which forms $2 \times 1024 \times 2$ images by assigning the time samples and DOST processed samples to different channels of its output image. This final image is then given as the input to the first layer of the trained convolutional neural network to perform the classification.

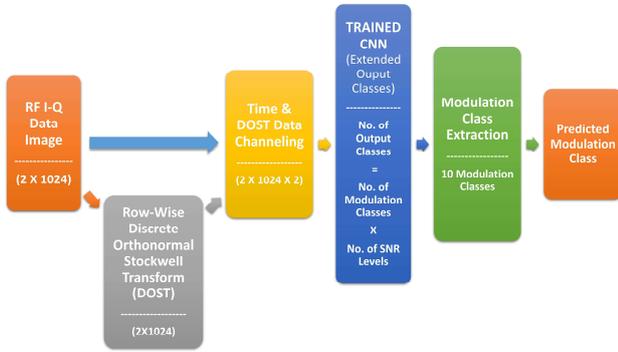


Fig. 1. Proposed System Model of Using Time and DOST Data Channeled Images and Extended Output Classes for CNN-based Radio Modulation Classification.

Before training the CNN, the entire pre-processing is done for all the training and validation I-Q time images as well, to convert them into their equivalent time and DOST channeled images. A detailed explanation of the architecture is presented in sub-section C of this section. The output layer of the CNN predicts one among the extended classes which are labelled according to both modulations as well as the SNR levels considered. This prediction of the extended class is then further processed in the Modulation Class Extraction block to extract the final modulation label for the input test data.

B. Time and Discrete Orthonormal Stockwell Transform Domain Channeling for RF I-Q Images

As described in the system model in sub-section A of this section, a new kind of pre-processing is incorporated in our approach to make the deep convolutional network learn signal features more prominently and efficiently. The output image of the DOST block contains all the time-frequency information from which the network can learn more features.

The physical intuition behind such pre-processing is that missing out on either the time data or time-frequency data could lead to loss of important features that the network could have learnt from a union of information of both the domains provided in a compact form. This led to the idea of creating two different channels in the same

image, similar to RGB channels in a digital photograph. One channel holds the data from time domain I-Q samples and the other channel holds the data from the DOST processed time-frequency domain image. This gives the final image a depth of 2, and the output size is $2 \times 1024 \times 2$. Moreover when the different kernels in a convolutional layer of the CNN train over such images, some of them might adapt to learn time features while some might adapt to learn time-frequency features. This was the inspiration behind the core idea of channeling multiple domain data. An example of the Time and DOST Domain Channeling for 16 I-Q time samples is shown below in Fig. 2.

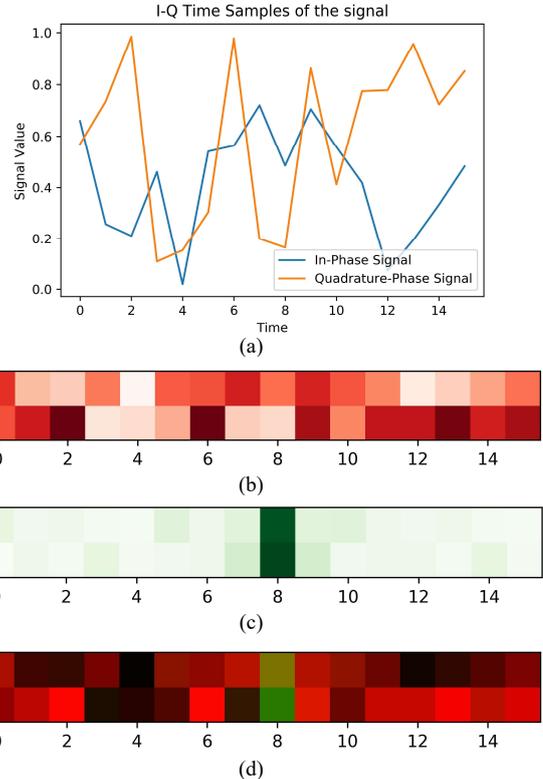


Fig. 2. (a) Plots of In-phase and Quadrature-phase time samples of the signal. (b) Visual representation of the RF I-Q image as the Red-Channel. (c) Visual representation of output of the DOST block as the Green-Channel. (d) Visual representation of the Time and DOST Domain Channeled Image in the form of an RGB image.

(*Note: White represents zero value in (b) and (c), but black represents zero value for RGB in (d) i.e. White in (a) + White in (b) = Black in (d))

Fig. 2(d) is just an equivalent representation of the Time-DOST Channeled image shown in a RGB format, and all the values in the blue channel are assumed zeros.

C. The CNN Architecture and Extended Output Classes

An eight layer convolutional neural network is proposed for the classification task in this work. It consists of 5 convolutional layers and 3 fully connected dense layers (including the output layer). The input image of size $2 \times 1024 \times 2$ is fed to the first convolutional layer (Conv) which has 128 filters, each of size 2×5 . The activation function used is ReLU, and appropriate zero padding is done to keep the output of the first layer the same size as that of the input image. The second, third and fourth convolutional layers are identical to the first layer. The fifth convolutional layer is different from the other four only with respect to the filter size i.e. 2×7 . The sixth and the seventh layers are fully connected dense layers (FC Dense)

with 256 neurons each and activation function used is ReLU. The eighth layer is the output layer, a fully connected dense layer with number of neurons equal to number of output classes, and SoftMax is used as the activation function. Here 10 modulations are considered for classification with the SNR levels varying from -8 dB to +8 dB in increments of 2 dB. Hence the number of extended output classes is 90 (No. of modulation classes multiplied by the No. of SNR levels).

Average pooling of pool size 1x4 is performed after the 1st and 2nd convolutional layer, and a pool size of 1x2 is used after the 3rd and 4th layers respectively. No pooling is performed after the 5th convolutional layer. The CNN layout is presented in Table I. Total number of trainable parameters is 1,861,850.

TABLE I. CNN ARCHITECTURE LAYOUT

Layers	Output Shape	Parameters
Input	2 x 1024 x 2	-
Conv 1 (128x2x5), ReLU	2 x 1024 x 128	2,688
Average Pooling (1x4)	2 x 256 x 128	-
Conv 2 (128x2x5), ReLU	2 x 256 x 128	163,968
Average Pooling (1x4)	2 x 64 x 128	-
Conv 3 (128x2x5), ReLU	2 x 64 x 128	163,968
Average Pooling (1x2)	2 x 32 x 128	-
Conv 4 (128x2x5), ReLU	2 x 32 x 128	163,968
Average Pooling (1x2)	2 x 16 x 128	-
Conv 5 (128x2x7), ReLU	2 x 16 x 128	229,504
FC Dense 6 (256), ReLU	256	1,048,832
FC Dense 7 (256), ReLU	256	65,792
FC Dense 8 (90), SoftMax	90	23,130

The filters used in the convolutional layers are of sizes 2x5 and 2x7. The main motive behind using 2-D filters is to allow the kernels to adapt to I and Q data separately.

The dataset used in this approach is a part of the RadioML 2018.01A real world, over the air captured dataset provided by DeepSig Inc. [21]. 10 primary modulations have been extracted from amongst the 24 modulation dataset for received SNR values ranging from -8 dB to 8dB with increments of 2 dB. These modulation classes include BPSK, QPSK, 8PSK, GMSK, 16APSK, 64 APSK, 16QAM, 64QAM, AM-DSB-WC and FM. We shall refer to this dataset as ‘Master Dataset’ in the rest of this article. The reason for choosing these specific modulation schemes is their extent of application in modern day communication systems like broadcast radios, satellite communication, satellite television, WLAN standards, Wi-MAX standards and cellular standards.

The concept of ‘Extended Output Classes’ is inspired from transfer learning. Each sample in the considered dataset is labelled with two tags i.e. the modulation tag and received SNR level tag. The general approach of using a CNN or any other deep neural network architecture for modulation classification on such a dataset has been classifying the input sample according to the modulation classes only. This has been presented in works [10] [11] and [9]. But in this proposed ‘Extended Output Classes’ method, the CNN is trained to predict both the modulation tag as well as the SNR tag of the input sample. This is done

by defining output classes with [Modulation, SNR] labels rather than just [Modulation] labels. For example, if the number of modulation classes is ‘M’ and the number of SNR levels considered in the dataset is ‘N’, then the number of extended output classes would be a product of M and N. The number of classes is increased by a factor N as compared to the general approaches.

In this work, data samples of 10 modulations over 9 SNR levels (-8:2:8 dB) are considered, hence instead of 10 output classes/modulation labels, we have 90 extended output classes which are the [Modulations, SNR] labels. Hence the last fully connected dense layer has 90 neurons, as presented in Table I. The main idea behind classifying on an extended class size is to make the network understand signal features at different SNR levels in a more adaptable manner and to prepare it for the varying SNR scenarios that it might face during testing on an unknown sample. For this the network should first learn to recognize the approximate SNR scenario from the input sample, and then adapt itself accordingly for achieving a better overall classification accuracy. An example of this process is illustrated in Fig. 3, which shows the extended [Modulation, SNR] output classes for a single modulation class ‘BPSK’.

Finally, as the last part of the system model, a modulation extraction block is used to extract only the [Modulation] label from the predicted [Modulation, SNR] labels by the CNN. The output of modulation extraction is one among the 10 classes of modulations that have been considered. This block can be implemented using a simple many-to-one mapping function.

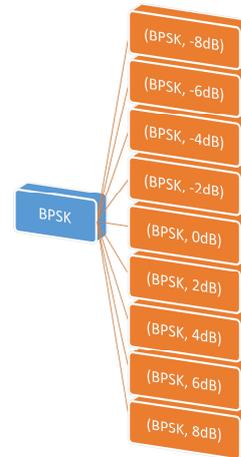


Fig. 3. Extended Output [Modulation, SNR] Classes shown for a single modulation class ‘BPSK’.

V. RESULTS AND ANALYSIS

The model for the CNN architecture described in subsection C of section IV was first built in using Keras. It was then trained, validated and tested on the Master Dataset. It contains a total of 368,640 samples, each sample being a 2x1024 RF I-Q Image in our considered dataset. 85% of the data samples are considered for the training and validation set, i.e. 313,344 samples, out of which 250,675 samples belong to the training set, and 62,669 samples belong to the validation set. The rest 55,296 samples are used as the test set. Training is performed using a categorical cross entropy cost function and an Adam optimizer.

We implement the training and prediction of our network in Keras [22] running on top of TensorFlow on a NVIDIA Cuda powered TESLA V100 16GB GPU in a Google Cloud Compute Engine Virtual Machine (VM) Instance. The VM instance was powered by a quad-core Intel Skylake based processor and 32GB of RAM.

The network was trained and evaluated for four different cases to test the effectiveness of Time & DOST Domain Channeling (T-D-D-C) and Extended Output Classes (E-O-C). Henceforward these two abbreviations shall be used in this article. In the first case, neither T-D-D-C was performed on the data samples, nor were E-O-C used. Hence the neuron count in the final FC dense layer falls to 10 from the previous count of 90. In the second case, T-D-D-C was not performed on the data samples, but E-O-C were used. In the third case, T-D-D-C was performed on the samples, but E-O-C were not used. Hence the network, in this case, is similar to that of the first case, with 10 neurons in the final FC dense layer. In the fourth and the final deciding case, T-D-D-C was performed on that data samples, as well as E-O-C were used.

The classification accuracies achieved by the network for all the four cases, over all the 10 different modulations and for different values of received SNR levels are plotted in Fig. 4. The average overall classification accuracies achieved by the network for all the four cases, over all the 10 different modulations and all values of received SNR levels are shown in Fig. 5.

In the first case, which can be considered to be the baseline case, the network achieved a maximum of 80.49% classification accuracy at 8 dB SNR and overall accuracy of 64.47%. This can be considered to be a very average classification performance.

In the second case the network achieved a maximum of 90.41% classification accuracy at 8 dB SNR and an overall accuracy of 71.82%. This can be considered to be a good push to the CNN performance. As it can be observed from Fig. 4, in this case the accuracies for SNR levels greater than and equal to 0 dB have been boosted, while those for SNR levels below 0 dB have remained more or less similar. Thus it can be inferred that by using extension of output classes, the network seems to have learnt to distinguish between good and bad SNR scenarios by just observing the data samples, and hence the training process has taken place accordingly to optimize performance.

In the third case the network achieved a maximum of 88.07% classification accuracy at 8 dB SNR and overall accuracy of 73.89%. As it can be observed from Fig. 4, in this case, although the maximum achieved accuracy is less as compared to that of the second case, the overall accuracy is more. Moreover the classification accuracies at all the SNR levels seem to have been boosted as compared to that of the first case. Also for SNR levels less than 2 dB, the classification accuracies are more as compared to that of the second case. Hence it can be inferred that by using time & DOST domain channeling on the data samples, the CNN was able to extract and learn more features from both the time domain samples as well as the time-frequency domain samples which accentuated its overall performance.

As a comparison with related papers, the results achieved in the second and third case in this paper already seems to outperform some particular results presented in

[11]. It can be observed by comparing our results with Fig. 7 and Fig. 8 of [11], that for almost the same number of training samples i.e. 250,675 in our case and 240,000 in [11], 10 modulations classes in our case as compared to 11 classes in [11], and 1024 I-Q time samples in a single RF I-Q image in both works, our second and third case already perform much better with respect to classification accuracies as well as understanding SNR scenarios. Also, the maximum SNR considered here is just 8 dB as compared to 20 dB in [11]. This comparison is done while our best case i.e. the fourth case/ proposed method is yet to be analyzed. This is done with an intention to show the novelty and effectiveness of E-O-C and T-D-D-C as independent methods to boost modulation classification performance.

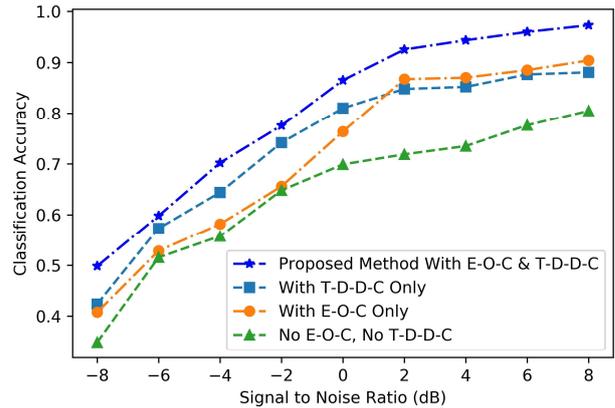


Fig. 4. Classification Accuracy vs. SNR for all the four cases of evaluation.

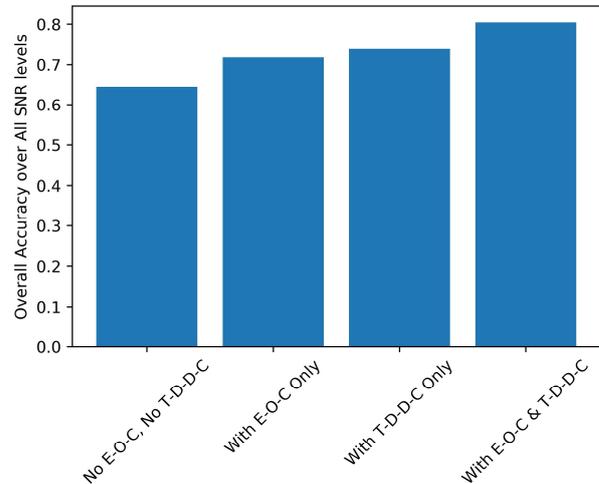


Fig. 5. Overall Classification Accuracy for all the four cases of evaluation.

In the fourth case the network achieved a maximum of 97.30% classification accuracy at 8 dB SNR and an overall accuracy of 80.45%. As it can be observed from Fig. 4, the performance of the CNN, in this case, is far superior to that of all the other cases considered. T-D-D-C and E-O-C combine together to form a strong and effective method to make the CNN learn more time and time-frequency features strongly as well as help it to understand the quality of signals and adapt the optimization process according to varying SNR scenarios. The CNN training process took 84 epochs, each lasting 45 seconds, with a batch size of 512 samples. The validation loss was monitored during the training.

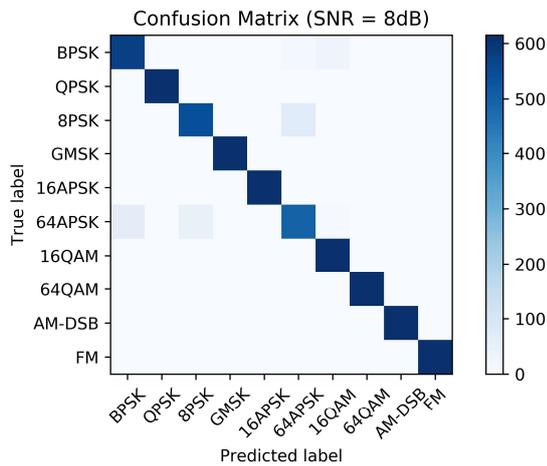


Fig. 6. Confusion Matrix for 10 modulation classes at 8 dB SNR.

To analyze the classification performance with respect to each individual modulation class, the confusion matrix at 8 dB SNR is shown in Fig. 6. It is observed from the figure that a nearly clean diagonal matrix is obtained at 8 dB SNR. The network is able to identify all modulation classes separately with very high accuracy. Though there is some slight confusion between the classes 8PSK and 64APSK, this might be because some of the constellation points are common to both the modulation schemes.

VI. CONCLUSION AND FUTURE SCOPE

Deep learning has seen a lot of development since the last decade. It has had unprecedented success in field like image classification, object recognition, natural language processing, unmanned vehicles, data analytics, and artificial intelligence. But its application in communication systems and devices is yet to be fully explored. This paper presents a way to use CNNs for the task of modulation recognition and also suggests a novel pre-processing method to improve upon their performance. This work might find scope in dynamic spectrum access and spectrum monitoring applications. As a continuation of this work in the future, other deep architectures like RNNs, LSTMs, ResNet structures for CNNs and fusion models as well as concepts of unsupervised learning can be explored in the domain of radio signal classification and identification. Also ensemble models trained individually for encountering samples at different SNR values can be explored to improve classification accuracy at low SNR scenarios. New pre-processing methods like T-D-D-C and suitable changes in the network architecture like E-O-C proposed in this paper can be further explored to accentuate performances of deep networks in specific scenarios. Hardware implementation solutions like FPGA acceleration of CNN as well as acceleration on dedicated low power edge computing devices can also be explored.

REFERENCES

- [1] O. A. Dobre, A. Abdi, Y. Bar-Ness and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET communications*, vol. 1, pp. 137-156, 2007.
- [2] P.-N. Tan and others, Introduction to data mining, Pearson Education India, 2007.
- [3] P. Brémaud, Markov chains: Gibbs fields, Monte Carlo simulation, and queues, vol. 31, Springer Science & Business Media, 2013.
- [4] W. Su, J. L. Xu and M. Zhou, "Real-time modulation classification based on maximum likelihood," *IEEE Communications Letters*, vol. 12, 2008.
- [5] J. L. Xu, W. Su and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, pp. 455-469, 2011.
- [6] P. Panagiotou, A. Anastopoulos and A. Polydoros, "Likelihood ratio tests for modulation classification," in *MILCOM 2000. 21st Century Military Communications Conference Proceedings*, 2000.
- [7] W. A. Gardner and C. M. Spooner, "Signal interception: performance advantages of cyclic-feature detectors," *IEEE Transactions on Communications*, vol. 40, pp. 149-159, 1992.
- [8] F. C. B. F. Muller, C. Cardoso Jr and A. Klautau, "A front end for discriminative learning in automatic modulation classification," *IEEE Communications Letters*, vol. 15, pp. 443-445, 2011.
- [9] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *Dynamic Spectrum Access Networks (DySPAN), 2017 IEEE International Symposium on*, 2017.
- [10] T. J. O'Shea, J. Corgan and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*, 2016.
- [11] T. J. O'Shea, T. Roy and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, pp. 168-179, 2018.
- [12] S. Hu, Y. Pei, P. P. Liang and Y.-C. Liang, "Robust Modulation Classification Under Uncertain Noise Conditions Using Recurrent Neural Networks," *IEEE Global Communications Conference (GLOBECOM)*, 2018.
- [13] D. Zhang, W. Ding, B. Zhang, C. Xie, C. Liu, J. Han and H. Li, "Heterogeneous Deep Model Fusion for Automatic Modulation Classification," *Preprints*, 1 2018.
- [14] Y. Wang and J. Orchard, "Fast discrete orthonormal Stockwell transform," *SIAM Journal on Scientific Computing*, vol. 31, pp. 4000-4012, 2009.
- [15] R. G. Stockwell, L. Mansinha and R. P. Lowe, "Localization of the complex spectrum: the S transform," *IEEE transactions on signal processing*, vol. 44, pp. 998-1001, 1996.
- [16] I. Daubechies and others, "Ten lectures on wavelets," in *CBMS-NSF regional conference series in applied mathematics*, 1991.
- [17] M. Eramian, R. A. Schincariol, R. G. Stockwell, R. P. Lowe and L. Mansinha, "Review of applications of 1D and 2D S transforms," in *Wavelet Applications IV*, 1997.
- [18] L. Mansinha, R. G. Stockwell, R. P. Lowe, M. Eramian and R. A. Schincariol, "Local S-spectrum analysis of 1-D and 2-D data," *Physics of the Earth and Planetary Interiors*, vol. 103, pp. 329-336, 1997.
- [19] L. Mansinha, R. G. Stockwell and R. P. Lowe, "Pattern analysis with two-dimensional spectral localisation: Applications of two-dimensional S transforms," *Physica A: Statistical Mechanics and its Applications*, vol. 239, pp. 286-295, 1997.
- [20] R. G. Stockwell, "A basis for efficient representation of the S-transform," *Digital Signal Processing*, vol. 17, pp. 371-393, 2007.
- [21] DeepSig, "RF Datasets for Machine Learning," DeepSig Inc., [Online]. Available: <https://www.deepsig.io/datasets/>.
- [22] F. Chollet, "Keras: The Python Deep Learning library," Keras, 2015. [Online]. Available: <https://keras.io/>.