Deep Learning Based Modulation Classification

Thesis submitted to the

Indian Institute of Technology Bhubaneswar

For award of the degree

of

Bachelor of Technology

by

Yenni Kaushik

Under the guidance of

Dr. Barathram Ramkumar

SCHOOL OF ELECTRICAL SCIENCES

INDIAN INSTITUTE OF TECHNOLOGY BHUBANESWAR

May 2020

CERTIFICATE

This is to certify that the thesis entitled **Deep Learning Based Modulation Classification**, submitted by **Yenni Kaushik** to Indian Institute of Technology Bhubaneswar, is a record of bonafide research work under my supervision and I consider it worthy of consideration for the award of the degree of Bachelor of Technology of the Institute.

Date :

Dr. Barathram Ramkumar
Associate Professor
School of Electrical Sciences
Indian Institute of Technology Bhubaneswar
Bhubaneswar, India

DECLARATION

I certify that

a. the work contained in the thesis is original and has been done by myself under the general supervision of my supervisor.

b. the work has not been submitted to any other institute for any degree or diploma.

c. I have followed the guidelines provided by the institute in writing the thesis.

d. I have conformed to the norms and guidelines given in the ethical code of conduct of the institute.

e. whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references.

f. whenever I have quoted written materials from other sources, I have put them under quotation marks and given due credit to the sources by citing them and giving required details in the references.

Yenni Kaushik

Acknowledgments

I would like to give my special thanks to Dr Barathram Ramkumar. His motivations and encouragements immensely enlightened me throughout my research. It has been a pleasure to perform research under his supervision. I appreciate his guidance, instructions and continuous support.

Yenni Kaushik

Abstract

In this work, we investigate the value of employing deep learning for the task of wireless signal modulation classification. We generate two data sets which simulate an AWGN channel and a Rayleigh channel. We consider a baseline method using cumulants and compare it with the deep learning approach across a varying range of signal-to-noise ratios . We use a convolutional neural network (CNN) architecture, a recurrent neural network (RNN) architecture and a Convolutional Long Short-term Deep Neural Network (CLDNN) architecure for the purpose of classification. Finally we conclude with an evalution of the performance of the architectures on the RadioML 2016 dataset.

Contents

List of Figures

List of Tables

List of Abbreviations

AMC	Automatic modulation classification
ANN	Artificial neural network
AWGN	Additive white Gaussian noise
CNN	Convolutional neural network
FB	Feature based
LB	Likelihood based
LSTM	Long short-term memory
CLDNN	Convolutional Long short-term deep neural network

List of Symbols

$\lvert \cdot \rvert$	Absolute value
$\mathbb{E}[\cdot]$	Expectation operator
$\exp(\cdot)$	Exponential

Chapter 1

INTRODUCTION

1.1 Motivation

Wireless communication plays an important role in modern communication. Therefore, modulation classification as an intermediate phase between signal detection and demodulation is gaining interest. Modulation recognition finds applications in commercial areas such as space communication and cellular telecommunication in the form of Software Defined Radios (SDR). SDR uses blind modulation recognition schemes to reconfigure the system, reducing the overhead by increasing transmission efficiency. Furthermore, AMC serves an important role in the information context of a military field. The spectrum of transmitted signals spans a large range and the format of the modulation algorithm varies according to the carrier frequency. The detector needs to distinguish the source, property and content correctly to make the right processing decision without much prior information. Under such conditions, advanced automatic signal processing and demodulation techniques are required as a major task of intelligent communication systems. The modulation recognition system essentially consists of three steps: signal preprocessing, feature extraction and selection of modulation algorithm. The preprocessing may include estimating SNR and symbol period, noise reduction and symbol synchronization. Deep learning algorithms have performed outstanding capabilities in images and audio feature extraction in particular and supervised learning in general, so it naturally comes as a strong candidate for the modulation classification task. To give a comprehensive understanding of AMC using deep learning algorithms, this project applies several state-of-art neural network architectures on simulated signals to achieve high classification accuracy.

1.2 Literature Survey

Over the past few decades, wireless communication techniques have been continuously evolving with the development of modulation methods. Communication signals travel in space with different frequencies and modulation types. A modulation classification module in a receiver should be able to recognize the received signal's modulation type with no or minimum prior knowledge. In adaptive modulation systems, the demodulators can estimate the parameters used by senders from time to time. There are two general classes of recognition algorithms: likelihood-based (LB) and feature-based (FB). The parameters of interest could be the recognition time and classification accuracy.

1.2.1 Likelihood-Based Methods

The LB-AMC has been studied by many researchers based on the hypothesis testing method. It uses the probability density function of the observed wave conditioned on the intercepted signal to estimate the likelihood of each possible hypothesis. The optimal threshold is set to minimize the classification error in a Bayesian sense. Therefore, it is also called the likelihood ratio test (LRT), because it's a ratio between two likelihood functions. The receiver measures the observed value of the input signal, then calculates the likelihood value under each modulation hypothesis. The estimated modulation algorithm is finally decided by the probability density functions.

The average likelihood ratio test(ALRT) algorithm proposed by Kim in 1988 [1], which successfully distinguished between BPSK and QPSK, is the first LB algorithm based on Bayesian theory. The authors in [1] assumed that signal parameters such as SNR, the symbol rate and carrier frequency are available for the recognizer.

LB methods are developed on complete theoretical basis, therefore derive the theoretical curve of the recognition performance and guarantee optimal classification results with minimum Bayesian cost. So it provides an upper bound or works as a benchmark for theoretical performance that can verify the performance of other recognition methods. Besides, by considering noise when building tested statistical models, LB presents outstanding recognition capability in low SNR scenarios. The algorithm can also be further improved for non-perfect channels according to the integrity of the channel information. However, the weakness of the LB approach lies in its computational complexity which may make the classifier impractical. When the number of unknown variables increases, it is hard to find the exact likelihood function.

Besides, if the assumption of prior information is not satisfied, the LB approach performance would decline sharply when the parameters are not estimated correctly or the built model does not match the real channel characteristics.

1.2.2 Feature-Based Methods

A properly designed FB algorithm can show the same performance as the LB algorithm but suffers from less computation complexity. The FB method usually includes two stages: extracting features for data representation and the decision making, i.e. classifiers. The key features can be categorized as time domain features including instantaneous amplitude, phase and frequency, transform domain features such as wavelet transform or Fourier transform of the signals, higher order moments (HOMs) and higher order cumulants (HOCs). The classifiers or pattern recognition methods include artificial neural networks, unsupervised clustering techniques, SVM and decision trees.

In 1995, Azzouz and Nandi [2] used instantaneous carrier frequency, phase and amplitude as key features and ANN as classifier, and conducted the recognition of analogue and digital signal schemes, which was considered as a new start of FB methods. Their simulation results show that the overall success rate is over 96% at the SNR of 15 dB using an ANN algorithm. It is indicated in [2] that the amplitude in 2-ASK changes in two levels which is equal in magnitude but opposite in sign. So the variance of the absolute value of the normalized amplitude contains no information, whereas the same function for 4-ASK contains information. A threshold is set in the decision tree for that distinguishing statistic. The maximum of the discrete Fourier transform of the instantaneous amplitude is calculated to discriminate FSK and PSK/ASK, as for the former the amplitude has information whereas it does not have for the latter two. M-PSK and ASK are distinguished according to the variance of the absolute normalized phase as ASK does not have phase information. The classifier is again chosen to be a binary decision tree.

We note that FB methods outperform LB methods in terms of preprocessing and the generality. It is based on a simple theory and the performance remains robust even with little prior knowledge or low preprocessing accuracy. But it is vulnerable to noise and non-ideal channel conditions.

1.2.3 Artificial Neural Networks

Artificial Neural Networks(ANNs) have succeeded in many research areas and applications such as pattern recognition and signal processing. Different kinds of neural networks have been implemented on the second step of feature based pattern recognition, including probabilistic neural networks and the support vector machine. Single multi-layer perceptrons (MLP) have been wildly used as classifiers as reported by L. Mingquan et al. [3] and Mobasseri et al. [4]. Others also suggested using cascaded MLP in ANN [2], in which the output of the previous layers are fed into latter layers as input.

Given the same input features, the MLP ANN outperforms the decision tree method. Unlike LB and FB approaches, where the threshold for decision should be chosen manually, the threshold in neural networks could be decided automatically and adaptively. On the other hand, as many decision-theoretic algorithms presented, the probability of a correct decision on the modulation scheme depends on the sequence of the extracted key features. As can be seen that a different order of key feature application results in different success rates for the modulation type at the same SNR. The ANN algorithms deal with this uncertainty by considering all features simultaneously, so that the probability of the correct decision becomes stable.

Recently ANN has been studied and improved to present outstanding performance in classification with the development of big data and computation ability. A deeper neural network outperforms traditional ANN by learning features from multilevel nonlinear operations. The concept of DNN was firstly proposed by Hinton [5] in 2006, which refers to the machine learning process of obtaining a multilevel deep neural network by training sample data. Traditional ANNs randomly initialize the weights and the bias in the neural network usually leads to a local minimum value. Hinton et al. solved this problem by using an unsupervised pre-training method for initialization of the weights.

DNN is generally categorized as feed-forward deep networks, feed-back deep networks and bi-directional deep networks. Feed-forward deep networks typically include MLP and CNN . CNN is composed of multiple convolutional layers and each layer contains a convolutional function, a nonlinear transformation and down sampling. Convolutional kernels detect the specific features across the whole input image or signal and achieve the weight sharing, which significantly reduces the computation complexity.

Unlike the ANN used in the traditional AMC problem, the deep neural network extracts the features inside its structure, leaving little preprocessing work for the receiver. Traditional AMC algorithms including FB and LB methods were proposed and tested on theoretical mathematical models.

1.3 Organization

In this thesis, we use simulated data as training and testing samples, and the data generation is introduced in Chapter 2. We select the cumulant based approach described in Chapter 3 as a baseline method. This thesis proposes different modulation classifiers by applying different deep neural network architectures as discussed in Chapter 4. These classifiers are trained on the RadioML 2016 dataset and the results are discussed in Chapter 5.

Chapter 2

EXPERIMENTAL SETUP

2.1 Data Generation

The Communications Toolbox (MATLAB) is used to generate synthetic training data. Each training example is a sequence of 128 samples with real and imaginary parts. In this project, we work with 2 data sets. In the first data set, we simulate an Additive white Gaussian noise channel. In the second data set, we simulate a Rayleigh fading channel. 100000 training examples are generated for each modulation scheme in both the data-sets. The SNR of the samples is uniformly distributed from -10 dB to +20 dB, with a step size of 2 dB. There are a total of 4 modulation schemes used:

- BPSK
- PAM-4
- 8-PSK
- QAM-4

2.2 Environment

For the purpose of training the data, we use Google Colaboratory. It's a Jupyter notebook environment that runs on a Tesla K80 GPU. This helps us in training the data in a fast and efficient manner.

Chapter 3

CLASSIFICATION USING CUMULANTS

3.1 Cumulants

For a complex valued stationary random process $y(n))$, second-order moments can be defined in two different ways.

$$C_{20} = E\left[y^2(n)\right] \tag{3.1}$$

$$C_{21} = E\left[|y(n)|^2\right] \tag{3.2}$$

Similarly, fourth-order moments and cumulants can be written in three ways. Thus, fourth-order cumulants can be defined as

$$C_{40} = cum\left(y(n), y(n), y(n), y(n)\right) \tag{3.3}$$

$$C_{41} = cum\left(y(n), y(n), y(n), y^*(n)\right) \tag{3.4}$$

$$C_{42} = cum\left(y(n), y(n), y^*(n), y^*(n)\right) \tag{3.5}$$

For zero-mean random variables w, x, y, and z, the fourth-order cumulant can be written as

$$cum\left(wxyz\right) = E\left(wxyz\right) - E\left(wx\right)\left(yz\right) - E\left(wy\right)\left(xz\right) - E\left(wz\right)\left(xy\right) \tag{3.6}$$

We can use (3.6) to express C_{40}, C_{41}, or C_{42} in terms of the fourth-and second-order moments of $y\left(n\right)$ with the appropriate conjugations [6]. The cumulants can be estimated from the sample estimates of the corresponding moments. We assume that $y\left(n\right)$ is zero-mean. Sample estimates of the correlations are given by

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N}\left|y\left(n\right)\right|^{2} \tag{3.7}$$

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N}y^{2}\left(n\right) \tag{3.8}$$

This leads to the following estimates:

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N}y^{4}\left(n\right) - 3\hat{C}_{20}^{2} \tag{3.9}$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N}y^{3}\left(n\right)y^{*}\left(n\right) - 3\hat{C}_{20}\hat{C}_{21} \tag{3.10}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N}\left|y(n)\right|^{4} - \hat{C}_{20}^{2} - 2\hat{C}_{21}^{2} \tag{3.11}$$

In practice, we estimate the normalized cumulants

$$\hat{C}_{4k} = \hat{C}_{4k}/\hat{C}_{21}^{2} \qquad k = 0, 1, 2 \tag{3.12}$$

We define a four-class problem based on the modulations given by

$$\Omega_4 = \{BPSK,\ PAM-4,\ QAM-4,\ PSK-8\} \tag{3.13}$$

$\left|\hat{C}_{40}\right|$ is used to decide whether the constellation is real-valued (BPSK/PAM), circular (PSK), or rectangular (QAM) [7].

$$\left|\hat{C}_{40}\right| < 0.34 \Rightarrow PSK\,(8)$$
$$0.34 \leq \left|\hat{C}_{40}\right| < 1.02 \Rightarrow QAM\,(4)$$
$$1.02 \leq \left|\hat{C}_{40}\right| < 1.68 \Rightarrow PAM\,(4)$$
$$1.68 \leq \left|\hat{C}_{40}\right| \Rightarrow BPSK$$

3.2 Results

The accuracy achieved on AWGN data-set with this approach is 64% on 1000 samples. For the Rayleigh data-set, the accuracy achieved is 59.4% on 1000 samples. It is observed that the accuracy doesn't improve much on increasing the number of samples.

Channel Type	**N**	**Accuracy**
AWGN	1000	64%
	10000	64.2%
Rayleigh	1000	59.4%
	10000	59.7%

Table 3.1: Accuracy for the four-class problem using cumulant based approach

Chapter 4

NEURAL NETWORK ARCHITECTURES

Deep learning greatly increased the capacity for feature learning directly on raw high dimensional input data based on high level supervised objectives due to the newly found capacity for learning of very large neural network models with high numbers of free parameters [8].

This was made possible by the combination of strong regularization techniques, greatly improved methods for stochastic gradient descent (SGD), low-cost, high-performance graphics-card processing power, as well as combining of key neural network architecture innovations such as convolutional neural networks, and rectified linear units.

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically include a series of convolutional layers that convolve with a multiplication or other dot product. The activation function is usually a RELU layer, and is subsequently followed by additional convolutions like pooling layers, fully connected layers and normalization layers, cited as hidden layers because their inputs and outputs are masked by the activation function and final convolution. The final convolution, in turn, often involves backpropagation so as to more accurately weigh the end product.

4.1 CNN-2

4.1.1 Architecture

The 2 layered CNN architecture used is as follows:

- Input shape - (400000, 2, 128, 1)[100000 samples for each modulation type]
- 1st Convolutional Layer- 64 filters of size 2 x 4
- 2nd Convolutional Layer- 16 filters of size 1 x 4
- Dense layer of size 64
- Dense layer of size 16
- The output layer uses a Softmax activation function and all the intermediate layers use a ReLu activation function.

4.1.2 Results

Although this architecture is simple in complexity and computationally light, it performed better than the cumulant-based approach by obtaining an accuracy of 76.2% on the AWGN data-set and 63.5% on the Rayleigh data-set.

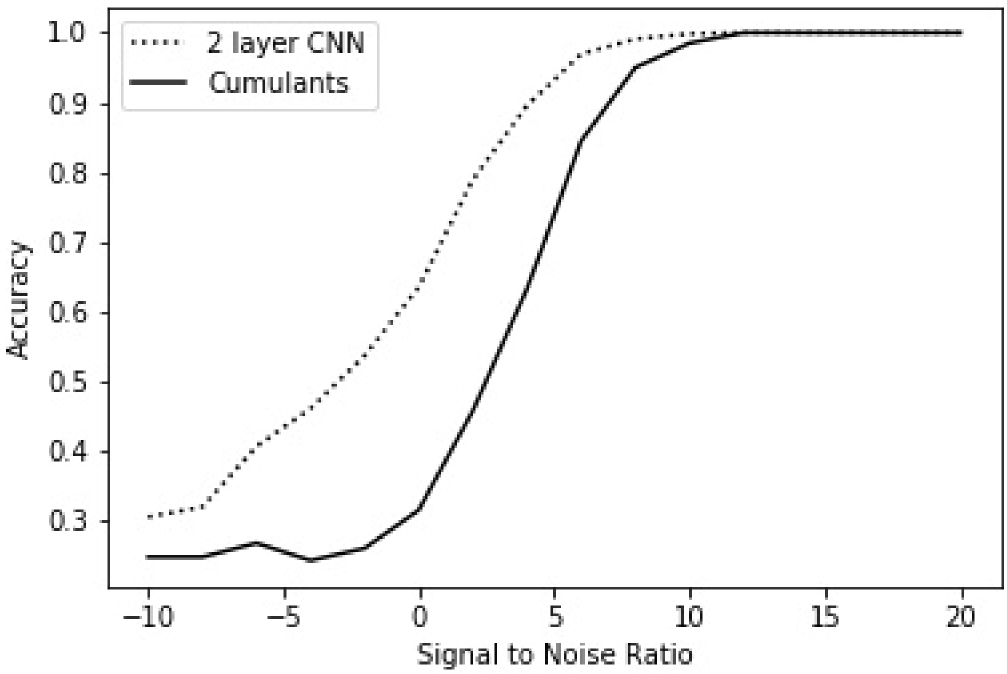

Figure 4.1: Comparison of classification accuracy between cumulants based approach and 2 layer CNN on the AWGN data-set with varying SNR

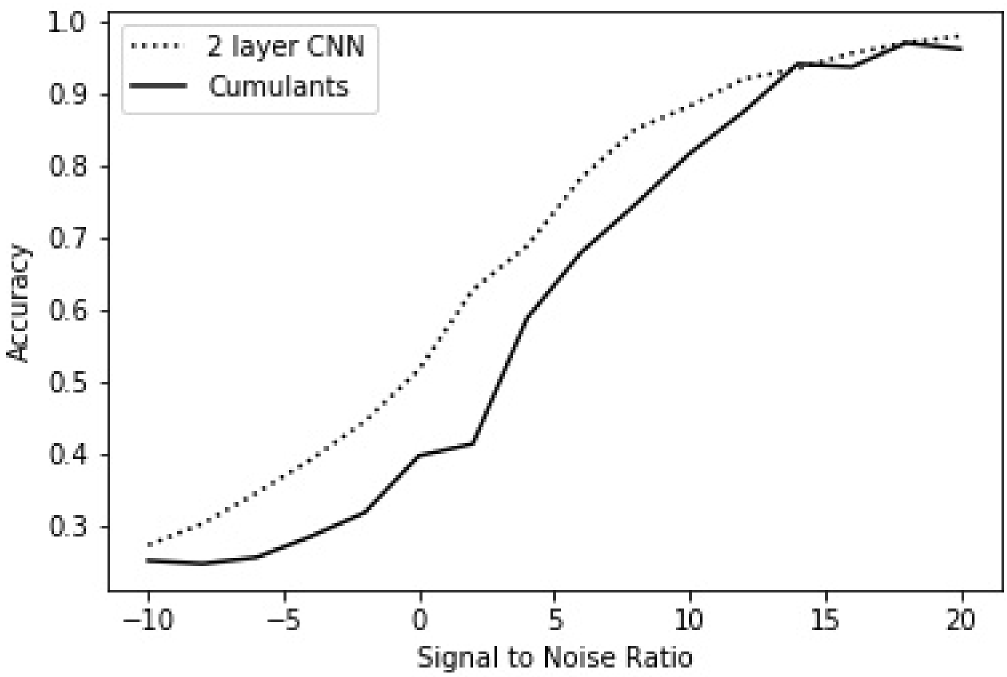

Figure 4.2: Comparison of classification accuracy between cumulants based approach and 2 layer CNN on the Rayleigh data-set with varying SNR

Channel Type	Accuracy
AWGN	76.2%
Rayleigh	63.5%

Table 4.1: Accuracy for the four-class problem using CNN with 2 layers

4.2 CNN-4

4.2.1 Architecture

We explore the optimal depth of CNN by increasing the number of convolutional layers from 2 to 4. We then explore the effect of different filter settings as in [9]. The architecture is as follows:

- Input shape - (400000, 2, 128, 1)[100000 samples for each modulation type]
- 1st Convolutional Layer- 256 filters of size 1x3
- 2nd Convolutional Layer- 256 filters of size 2x3
- 3rd Convolutional Layer- 80 filters of size 1x3
- 4th Convolutional Layer- 80 filters of size 1x3
- Dropout - 0.5
- Zero Padding of 2 units after every layer.
- Dense layer of size 256.
- The output layer uses a Softmax activation function and all the intermediate layers use a ReLu activation function.

4.2.2 Results

This architecture achieved an accuracy of 80.7% on the AWGN data-set and 68.3% on the Rayleigh data-set.

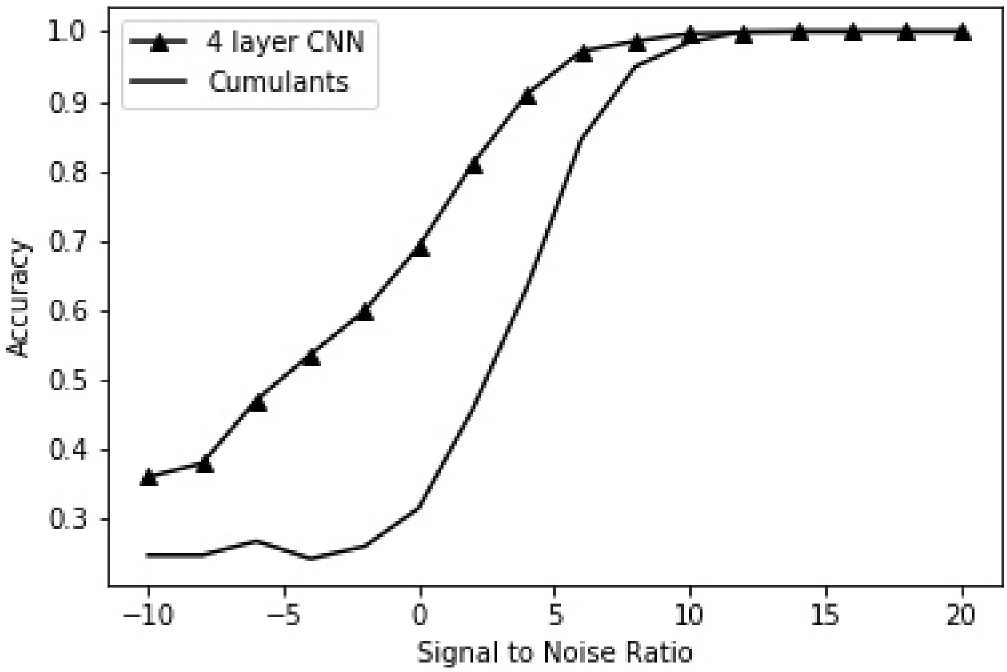

Figure 4.3: Comparison of classification accuracy between cumulants based approach and 4 layer CNN on the AWGN data-set with varying SNR

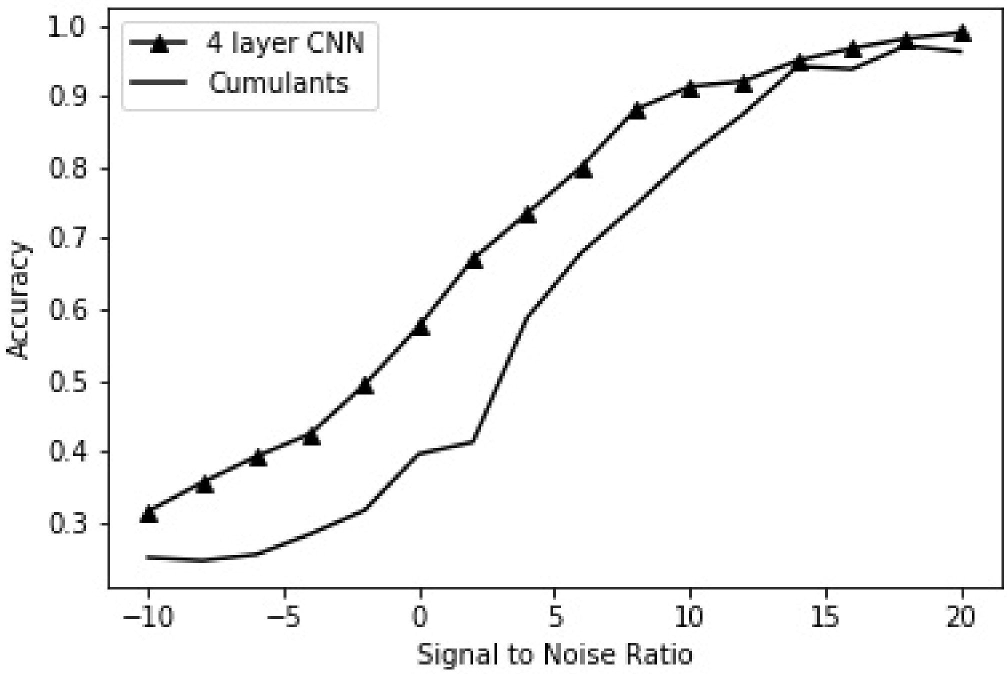

Figure 4.4: Comparison of classification accuracy between cumulants based approach and 4 layer CNN on the Rayleigh data-set with varying SNR

Channel Type	Accuracy
AWGN	80.7%
Rayleigh	68.3%

Table 4.2: Accuracy for the four-class problem using CNN with 4 layers

4.3 LSTM

4.3.1 Architecture

Conventional RNN suffers from vanishing gradient problem during the gradient descent updates. Hence, RNN is not suitable to learn the long-term dependencies. Gated RNN has been introduced to overcome such limitation.

LSTM neural network is one kind of gated RNN with LSTM units. The core idea behind LSTM unit is that besides the outer unit recurrence, it has LSTM cells that have an internal recurrence (self-loop). Each cell has the same inputs and outputs as an ordinary recurrent neural network, but introduces more parameters and a system of several gates that controls the flow of information [10].

The LSTM architecture used is as follows:

- Input shape - (400000, 128, 2)[100000 samples for each modulation type]
- LSTM layer of size 16
- Dense layer of size 16 with a ReLu activation function
- The output layer uses a Softmax activation function

4.3.2 Results

This architecture achieved an accuracy of 78% on the AWGN data-set and 66.7% on the Rayleigh data-set.

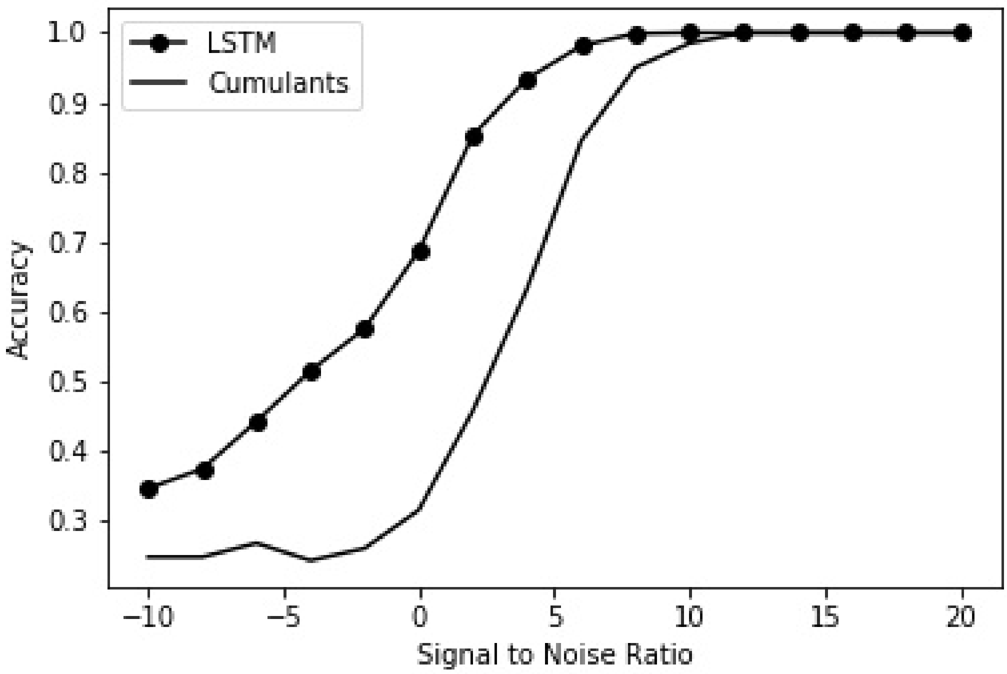

Figure 4.5: Comparison of classification accuracy between cumulants based approach and LSTM network on the AWGN data-set with varying SNR

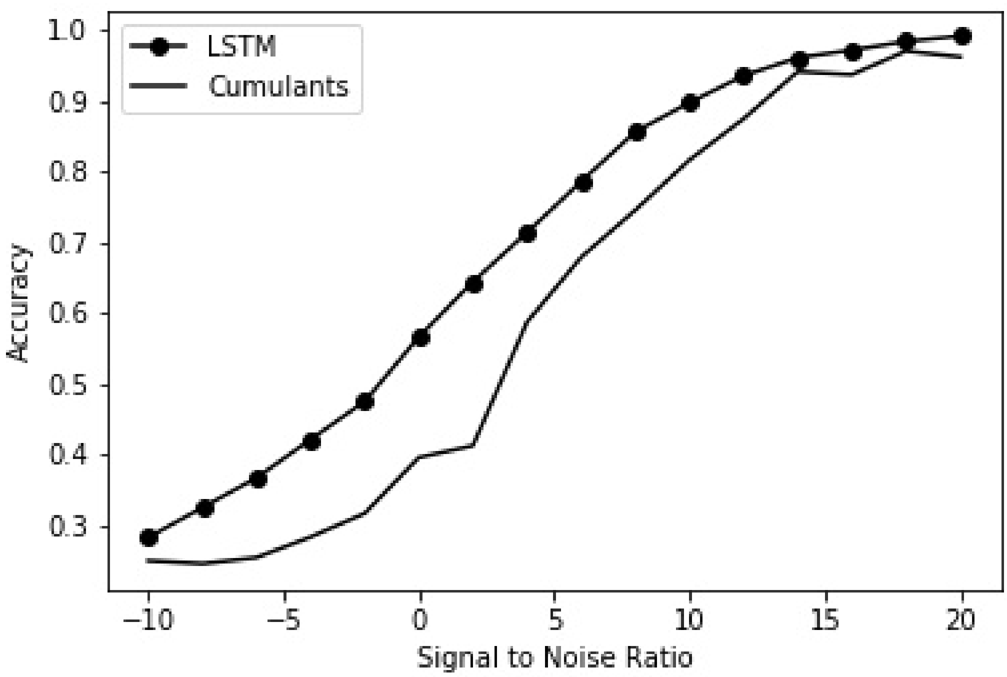

Figure 4.6: Comparison of classification accuracy between cumulants based approach and LSTM network on the Rayleigh data-set with varying SNR

Channel Type	**Accuracy**
AWGN	78%
Rayleigh	66.7%

Table 4.3: Accuracy for the four-class problem using LSTM neural network

4.4 CLDNN

4.4.1 Architecture

CLDNN has been widely used in recognition tasks that involve time domain signals like videos, speech, and images, as the inherent memory property leads to recognizing temporal correlations in the input signal. An LSTM unit is added into the network after the convolutional part to extract more relevant temporal features in the signal.

The CLDNN architecture used is as follows:

- Input shape - (400000, 2, 128, 1)[100000 samples for each modulation type]
- 1st Convolutional Layer- 256 filters of size 1x3
- 2nd Convolutional Layer- 256 filters of size 2x3
- 3rd Convolutional Layer- 80 filters of size 1x3
- 4th Convolutional Layer- 80 filters of size 1x3
- Dropout - 0.5
- Zero Padding of 2 units after every layer.
- LSTM layer with 50 computing units.
- Dense layer of size 256.

- The output layer uses a Softmax activation function and all the intermediate layers use a ReLu activation function.

4.4.2 Results

This architecture achieved an accuracy of 83.1% on the AWGN data-set and 70.1% on the Rayleigh data-set.

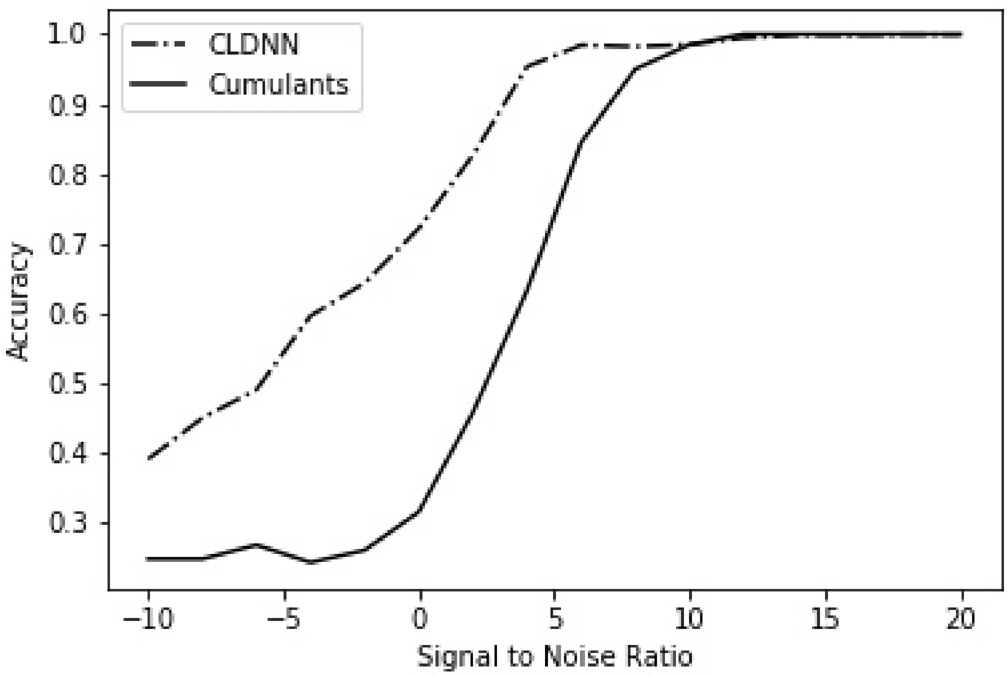

Figure 4.7: Comparison of classification accuracy between cumulants based approach and CLDNN on the AWGN data-set with varying SNR

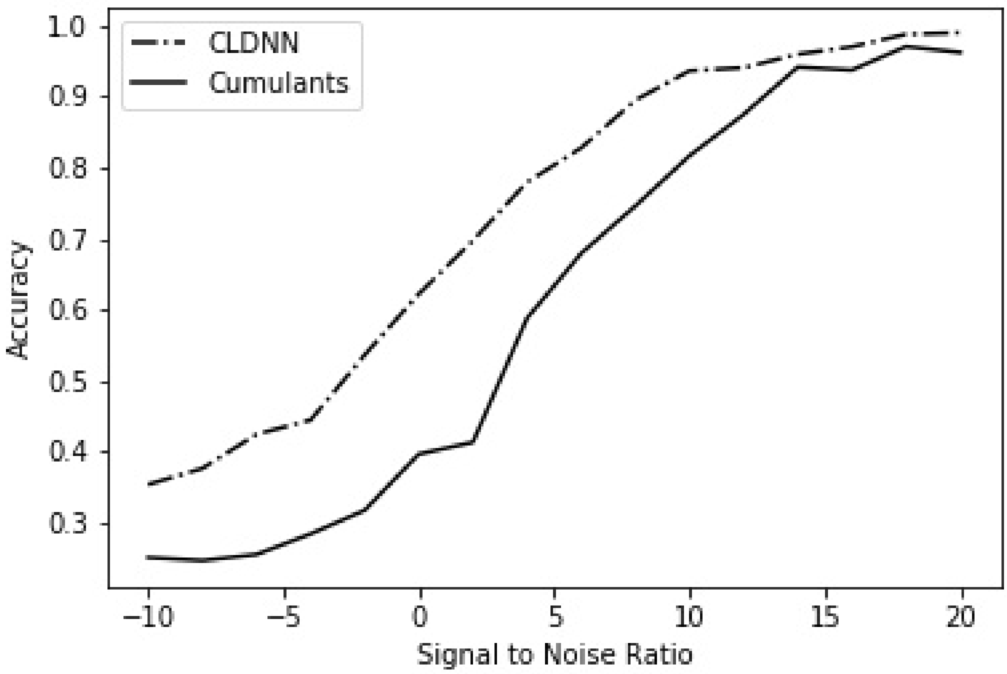

Figure 4.8: Comparison of classification accuracy between cumulants based approach and CLDNN on the Rayleigh data-set with varying SNR

Channel Type	**Accuracy**
AWGN	83.1%
Rayleigh	70.1%

Table 4.4: Accuracy for the four-class problem using CLDNN

Chapter 5

RadioML 2016

This is a synthetic dataset, generated with GNU Radio, consisting of 11 modulations (8 digital and 2 analog). This is a variable-SNR dataset with moderate LO drift, light fading, and numerous different labeled SNR increments for use in measuring performance across different signal and noise power scenarios. The data is generated in a way that captures various channel imperfections that are present in a real system using GNU radio.

The training examples – each consisting of 128 samples - are fed into the neural network in 2 x 128 vectors with real and imaginary parts separated in complex time samples. The labels in the input data include the SNR and the modulation type. The SNR of the samples is uniformly distributed from -20 dB to +18 dB, with a step size of 2 dB. The data set is split equally among all considered modulation types. These are the modulation types:

- 8-PSK
- BPSK
- CPFSK
- GFSK
- PAM-4
- QAM-16

- QAM-64
- AM-DSB
- AM-SSB
- WBFM

5.1 Results

For the purpose of evaluating the performance of various architectures, the data-set is split into train and test sets in the ratio of 9:1. After training each model with the training data, we compare their classification accuracies on the test data.

The CNN-2 (2 convolutional layers) architecture achieved an accuracy of 56.1% on the data set. The best accuracy achieved at high SNR was approximately 82%. The accuracy declined with the decline in SNR.

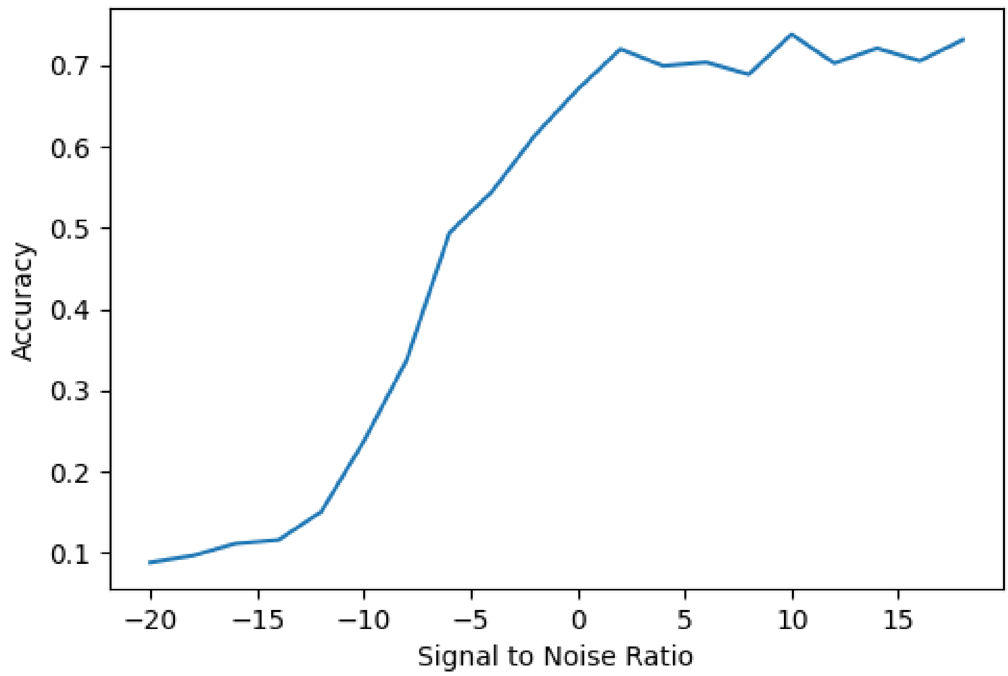

Figure 5.1: Classification Accuracy of CNN-2 architecture vs SNR

The CNN-4 (4 convolutional layers) architecture achieved an accuracy of 60.3% on the data set. The best accuracy achieved at high SNR was approximately 84%.

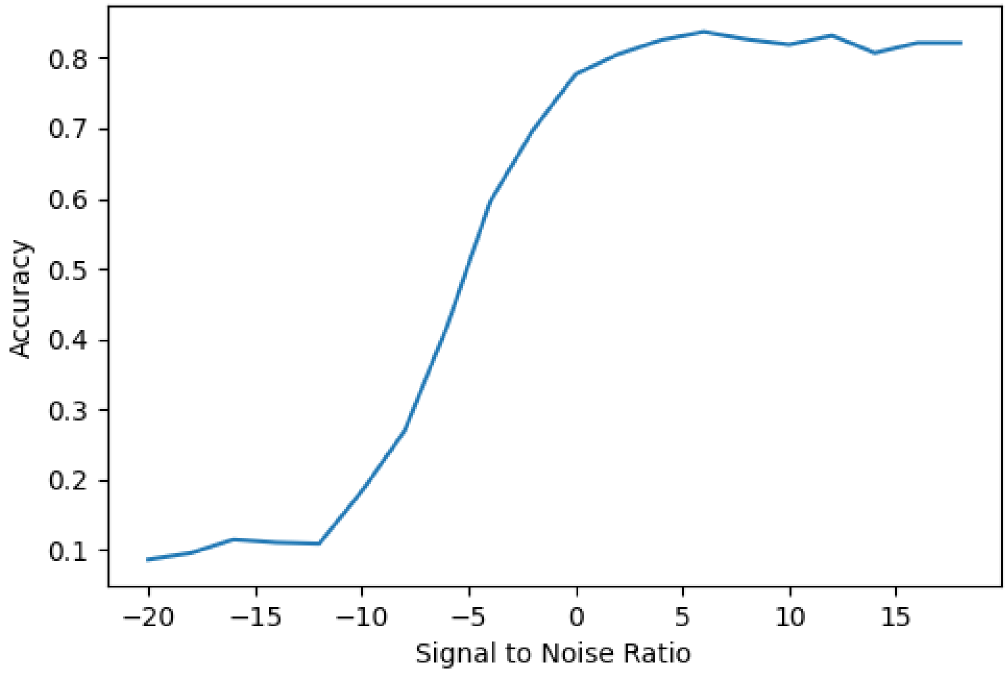

Figure 5.2: Classification Accuracy of CNN-4 architecture vs SNR

The LSTM (Long short-term memory) architecture achieved an accuracy of 54.5% on the data set. The best accuracy achieved at high SNR was approximately 82%.

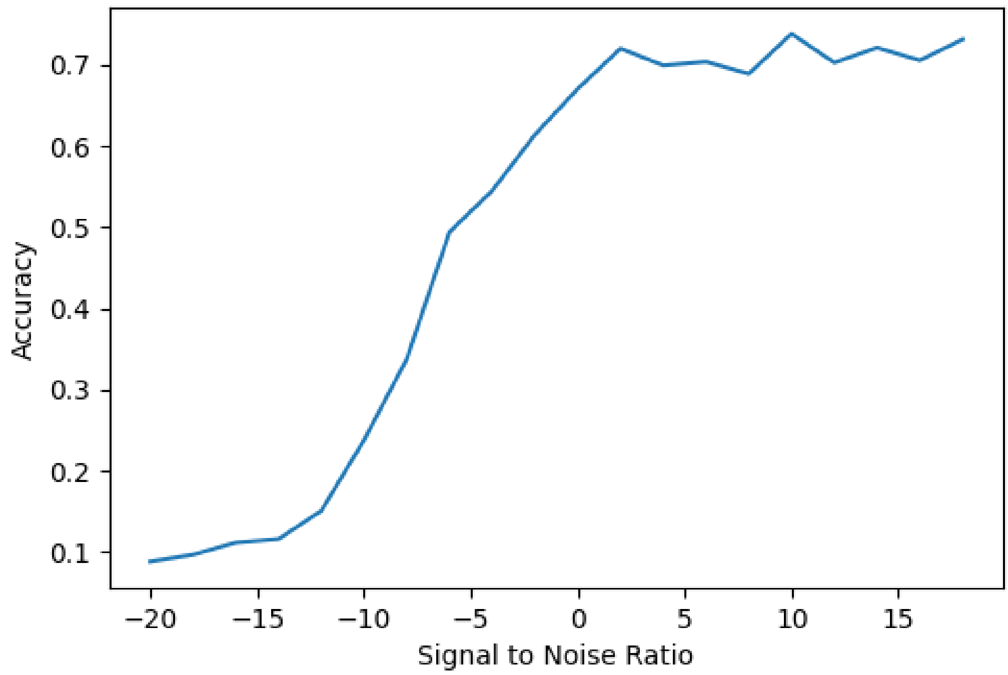

Figure 5.3: Classification Accuracy of LSTM architecture vs SNR

The Convolutional long short-term deep neural network architecture achieved an accuracy of 61.6% on the data set. The best accuracy achieved at high SNR was approximately 85%.

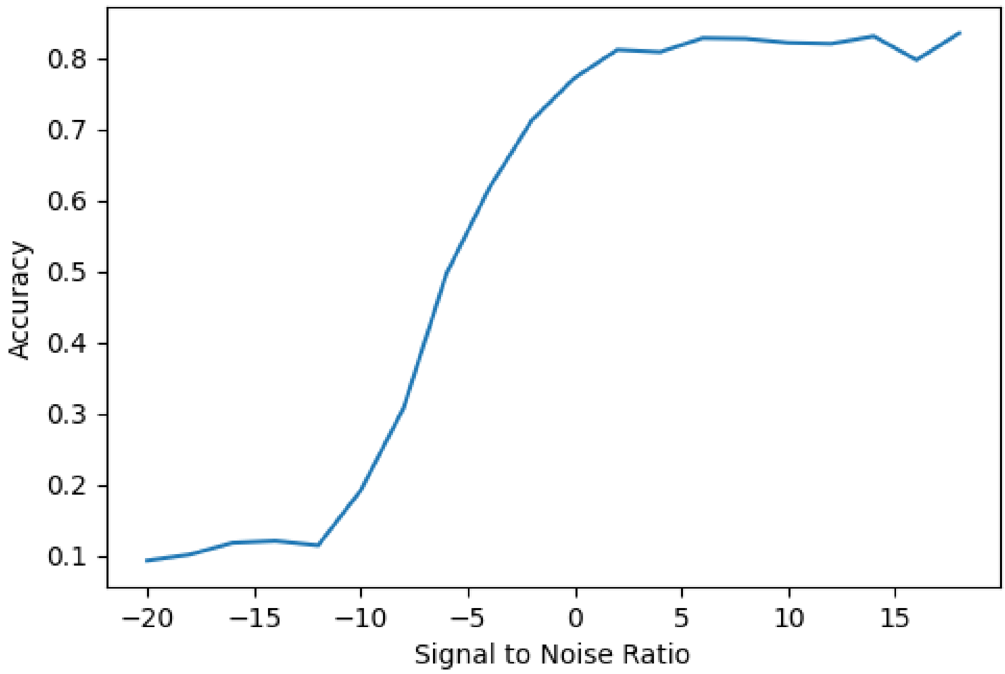

Figure 5.4: Classification Accuracy of CLDNN architecture vs SNR

Chapter 6

Conclusion

This thesis has implemented several deep learning neural network architectures for the automatic modulation classification task. Multiple classifiers are built and tested, which provide high probabilities of correct modulation recognition in a short observation time. The trained models outperform traditional classifiers by their high success rates and low computation complexities. The CNN serves as a basic end-to-end modulation recognition model providing nonlinear mapping and automatic feature extraction. A CLDNN model combines a CNN block, a LSTM block and a DNN block as a classifier that can automatically extract the spatial and temporal key features of signals. This model produces the highest accuracy for time domain IQ inputs. The experiments of time domain IQ and amplitude phase inputs also emphasize the importance of preprocessing and input representation. These models are capable of recognizing the modulation formats with various propagation characteristic, and show high real-time functionality.

Bibliography

[1] K. Kim and A. Polydoros, "Digital modulation classification: the bpsk versus qpsk case," in *MILCOM 88, 21st Century Military Communications-What's Possible?'. Conference record. Military Communications Conference.* IEEE, 1988, pp. 431–436.

[2] E. E. Azzouz and A. K. Nandi, "Automatic identification of digital modulation types," *Signal Processing*, vol. 47, no. 1, pp. 55–69, 1995.

[3] L. Mingquan, X. Xianci, and L. Lemin, "Ar modeling-based features extraction of multiple signals for modulation recognition," in *ICSP'98. 1998 Fourth International Conference on Signal Processing (Cat. No. 98TH8344)*, vol. 2. IEEE, 1998, pp. 1385–1388.

[4] B. G. Mobasseri, "Digital modulation classification using constellation shape," *Signal processing*, vol. 80, no. 2, pp. 251–277, 2000.

[5] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[6] D. R. Brillinger, *Time series: data analysis and theory.* Siam, 1981, vol. 36.

[7] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Transactions on communications*, vol. 48, no. 3, pp. 416–429, 2000.

[8] 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, no. 1, pp. 168–179, 2018.

[9] X. Liu, D. Yang, and A. El Gamal, "Deep neural network architectures for modulation classification," in *2017 51st Asilomar Conference on Signals, Systems, and Computers.* IEEE, 2017, pp. 915–919.

[10] S. Hu, Y. Pei, P. P. Liang, and Y.-C. Liang, "Robust modulation classification under uncertain noise condition using recurrent neural network," in *2018 IEEE Global Communications Conference (GLOBECOM).* IEEE, 2018, pp. 1–7.