Deep Learning based Automated Modulation Recognition for Cognitive Radio Networks

Syed Muhammad Shehram Shah†, Sajjad Ali Memon††, Erum Saba †††, Abdul Latif†† and Aqeel Ahmed††

†Department of Software Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan
††Department of Telecommunication Engineering, Mehran University of Engineering and Technology, Jamshoro, Pakistan
†††Information Technology Centre, Sindh Agriculture University, Tando Jam, Sindh, Pakistan

Summary
Cognitive radio communication systems are regarded as the future of wireless communication systems since they can cater to demands in terms of data rate, latency, and quality-of-service. One of the challenges for cognitive radios is the automated recognition of modulation schemes, which is required for identifying unknown secondary transmissions. Automated modulation recognition (AMR) is the task of recognizing the type of modulation scheme used by the unknown secondary transmission sources from amongst a pool of possible modulation schemes. In this paper, we benchmark the performance of two types of deep neural network (DNN) architectures against a feature-engineered machine learning model based on Cumulant features and Logistic Regression. We show that DNN architectures outperform the handcrafted features models, thereby, highlighting the well-known learning ability of DNNs. We also show that the ability of DNNs to recognize modulation schemes is limited by the SNR of the received modulation symbols.

Key words:
Automated Modulation Recognition; Cognitive Radio, Wireless Communication; Deep Learning

1. Introduction

The global demand for data communication is increasing courtesy of everygrowing mobile wireless communication infrastructure. Global mobile data consumption increased from 7.20 exabytes per month in 2016 to 19.01 exabytes per month by the end of 2018, i.e., an increase of 62% [Cisco2016, Cisco2016a]. With more applications on the way, including the internet of things, smart cities, e-health, e-learning etc., the demand for mobile data was expected to reach 28.56 exabytes per month by the start of 2020, and is expected to reach 56.80 exabytes per month by 2021. As a result, wireless communication systems now require regular technological evolution in order to improve their data-rate, latency, and quality-of-service. One may not be wrong to believe that the future of wireless communication systems essentially depends on their versatility, adaptability, intelligence, and in some sense cognitive ability. It has been argued that cognitive radios may be the answer to many such demands of future wireless communication systems [Gupta2015, Ejaz2016, Sexton2017].

The term cognitive radio was coined by Mitola et al. [Mitola1999], who defined it as a radio that employs model-based reasoning to achieve a specified level of competence in radio-related domains. Research into the field of cognitive radios has since received a lot of interest from the radio engineering community, mostly because these automated radio communication systems offer advantages in terms of intelligent networking [Lu2007], automated spectrum management [Niyato2008], improved spectrum access and capacity [Neel2007], as well as optimized radio link performance [Rondeau2004].

Cognitive radios have two main attributes, these include: a) the ability to sense, observe, and map the radio environment, and b) software-based reconfigurability of the physical layer [Haykin2005]. These two attributes are controlled by a cognition engine that enables the cognitive radio to build a radio environment map and then decide on the best possible transmission system to enable communication in that radio environment. The radio environment map is typically built by sensing the radio spectrum over time, space, and frequency, and subsequently cataloging the types of transmissions which are taking place over the wireless channel [Yucek2009]. An integral part of this radio environment mapping process is the automated recognition of modulation schemes which are used for these unknown transmission(s) [Dobre2007, Weber2015].

Automated modulation recognition (AMR) is the task of recognizing the type of modulation scheme used by the unknown transmission source(s) from amongst a pool of possible modulation schemes. The objective here is to recognize the modulation scheme without prior information about the channel conditions or the nature of transmission. To this end, AMR has become a popular research topic in radio communication with several researchers working towards the development of AMR methods.

In research literature, one finds that traditional AMR methods were based on either decision theory or feature

Manuscript received May 5, 2022
Manuscript revised May 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.5.56

engineering approach which employed hand-crafted features followed by a classifier [Xu2011, Park2008, Lv2014]. With the advent of the deep learning era, the concept of end to-end learning has become popular. Here, the idea is that raw data i.e., signals without pre-processing or feature-engineering, can be provided as input to deep neural networks which cannot only learn to extract relevant features in an automated manner but also optimize a classifier in order to maximize its ability to distinguish between various classes [LeCun2015]. In fact, end-to-end learning mechanisms have already proven their worth for tasks related to image processing [Zhang2016], speech processing [Zhang2017], and natural language processing [Devlin2018]. To this end, several researchers have also proposed end-to-end learning for AMR, where the objective is to train deep neural networks directly on modulated symbols in complex baseband representation [Jagannath2018, OShea2018].

In this paper, we benchmarked the classification performance of a VGGnet-style convolutional neural network (CNN) and a residual network against a model based on handcrafted cumulant features and logistic regression classifier. We employed these models to classify seven different digital modulation schemes. In order to train and test these models, we use a publicly available dataset which consists of transmissions of modulation schemes over various types of channels. The dataset is publicly available for academic research, making our research reproducible.

The rest of the paper is structured as follows; In Section 2, we review of important contributions in this domain is provided. In Section 3, we provide a brief summary of the RadioML 2018.01A dataset which is used in our experiments. Here, we also provide detailed information about the training and validation partitions. In Section 4, we provide insights into our experiments we detail the methods used for developing and training DNNs. In Section 5, we provide results of our experiments, both numerically and pictorially, and finally, close the paper with a conclusion in Section 6.

2. Related Work

As mentioned earlier, the success of deep learning in the fields of audio, visual, and textual computing has also motivated researchers to propose deep learning-based method for automated recognition of modulation schemes. To this end, we note that Nandi et al. [Nandi1998] were amongst the first to propose the use of artificial neural networks for the classification of analog and digital modulated signals for the communication system intelligence.

Amongst other works, one can argue that the work of O'Shea et al. [OShea2016] is most significant because they provided a publicly available dataset in the form of RadioML2016a which provided the biggest impetus to research towards AMR. Here, they proposed the use of convolutional neural networks for recognition of a set of modulation schemes that included 3 analog and 8 digital modulation schemes.

In [Hong2017] Hong et al. and in [West2017] West et al. proposed the use of recurrent neural networks (RNNs) for classification of modulation schemes. The most interesting aspect of the work from West et al. is that they reported that depth did not have a significant influence on the classification performance of their proposed architectures.

Liu et al. investigated the efficacy of various deep neural network architectures in [Liu2017] using the DeepSig RadioML2016.1b dataset. They reported that Convolutional Long Short-term Deep Neural Network (CLDNN) architecture [Sainath2015] achieved the best classification accuracy (88.5%) as compared to CNN (83.8%), ResNet (83.5%), and densely connected convolutional neural networks (DenseNet) [Huang2017] (86.6%). However, the results from their work should be interpreted with caution since they combined analogue and digital modulation schemes together in their dataset, even though analogue modulation schemes such as amplitude modulation, frequency modulation are not used for cognitive radio systems.

O'Shea et al. [OShea2018] published an extended and improved version of their dataset in 2018, which they called the RadioML2018.01a, similar to their previously published dataset i.e. the RadioML2016a [OShea2016]. Sabour et al. [Sabour2017] investigated Capsule networks which have recently gained traction for object recognition tasks. However, they report that Capsule networks took almost 32 hours to train as compared to CNNs, which took 7.5 minutes, or long-short term memory (LSTM), which took 63 minutes. Their work is important because it suggests the difficulty in training both recurrent and Capsule networks for the recognition of modulation schemes.

In [Zha2019], Zha et al. and in [Peng2019], Peng et al. proposed to transform modulated symbols from a time-series representation into images using either short-time Fourier transform or scatterplots of symbol constellation, and subsequently used image classifiers for achieving the task of AMR. One can argue, however, whether such transformations are viable for real-time applications, especially since one can develop deep learning architectures that can work directly with time-series signals.

3. Dataset

We used DeepSig RadioML2018.01a dataset, which was published by O'Shea et al. in [OShea2018]. The dataset consists of more than 2 million complex-valued symbols (in terms of amplitude values of I/Q channels) from 24 different analog and digital modulation schemes. Channel impairments are added to these symbols by emulating over-the-air transmission in ISM 900 MHz band and through synthetic multipath propagation in a Rayleigh fading environment.

We use a subset of the DeepSig RadioML 2018.01A dataset. To this end, we first eliminate all examples of analog modulation schemes, thus we are only left with examples from digital modulation schemes. The motivation is that one does not expect analog modulation schemes to be used in cognitive radios. Moreover, amongst digital modulation schemes, we also remove examples of schemes such as amplitude shift keying, frequency-shift keying, and minimum shift keying since these modulation schemes are not used in modern wireless communication systems. This leaves examples from seven modulation schemes in our subset of DeepSig RadioML 2018.01A dataset. These include BPSK, QPSK, 16-QAM, 32- QAM, 64-QAM, 128-QAM, and 256-QAM.

The subset of DeepSig RadioML 2018.01A consists of 576,000 examples in a total of these modulation schemes. We partition the subset into training and validation partition with an 80% and 20% training/validation ratio. The signal-to-noise ratio (SNR) for modulated symbols is varied between 0 dB and 30 dB, in incremental steps of 2 dB. For every value of SNR, we have 4000 examples for each of the seven modulation schemes.

4. Methods

In this section, we shall provide a brief description of the CNN and ResNet based DNNs used in our analysis along with the method for computing cumulant features.

5.1 Convolution Neural Network

The first architecture we shall investigate is based on Convolutional Neural Network (CNN), in particular the one based on the famous VGGNet model [Simonyan2015], which has performed exceedingly well for various object recognition tasks, is built for investigation. Our model has a slight modification to the original VGG-Net model, that is, we use batch normalization instead of dropout for regularization since it has been shown to be better than dropout for this purpose [Li2018].

We develop the CNN model in the form of blocks. Each block has three convolution layers. A block starts with an input layer that passes two-dimensional modulated symbols (as I/Q components) to the CNN. The first convolutional layer has 16 kernels of size (3,2), where the size (3,2) represents the dimension × (*I-channel*+*Q-channel*). This is followed by batch normalization and ReLU activation function. Next, a convolutional layer with 32 filters and size (3,1) is added to extract higher level features than the first convolution layer and is followed by ReLU, batch normalization, and a max-pooling layer of size (2,1) to reduce the dimensions of the activation map by half across the temporal axis. The final layer of our CNN block is a convolutional layer with 16 kernels size of (1,1). We use this layer to reduce the number of convolutional filters from 32 to 16. Now, to investigate model depth and complexity, the CNN block is repeated up to six times. Finally, a fully connected dense layer with seven outputs (corresponding to the number of classes) along with the sigmoid activation function is used for classification.

5.2 Residual Networks

In addition to the VGGNet inspired CNN, we also experiment with (Residual Networks (ResNets). Similar to the CNN, ResNet models are also built-in terms of blocks. A block consists of a convolutional first layer with 16 kernels of size (3,2). This is followed by batch normalization and ReLU activation function. The second convolutional layer consists of convolutional layer with 32 kernels of size (3,1) along with batch normalization and ReLU activation. To investigate the effect of model depth and complexity, we repeat this block up to six times (note that the kernel size is (3,1) from the second block onwards).

5.3 Benchmarking DNN models

To benchmark the performance of DNN models, we consider a feature-engineering based classification paradigm which includes training a logistic regression classifier with higher-order-statistical features based on cumulants. Cumulants are functions of descriptive statistics which provide information about the probability distribution of dataset, which in our case contains modulated symbols. In [Swami2000], Swami et al. hypothesized that the probability distribution of each type of modulation scheme is unique and can be modelled to classify between them. They proposed the use of 4^{th} order cumulant features for AMR. It is important to mention here that while their work highlighted the usability of cumulant based features, their classification methodology relied on analytically derived thresholds and subsequent comparison of cumulant features with these thresholds. Following [Wu2008, Chang2015]. work, the second, fourth, and sixth order cumulants for modulated symbols can be computed .

If s[n] represents the nth modulated symbol received amongst a total of N symbols, then cumulants of s[n] can be computed as:

Second order cumulants

$$C_{20} = M_{20} \tag{2}$$

$$C_{21} = M_{21} \tag{3}$$

Fourth order cumulants

$$C_{40} = M_{40} - 3M_{20}^{0} \tag{4}$$

$$C_{41} = M_{41} - 3M_{20}M_{21} \tag{5}$$

$$C_{42} = M_{42} - M_{20}^{2} - M_{21}^{2} \tag{6}$$

Sixth order cumulants

$$C_{60} = M_{60} - 15M_{20}M_{40} + 30M_{30}^{3} \tag{7}$$

$$C_{61} = M_{61} - 10M_{20}M_{41} - 5M_{21}M_{40} + 30M_{21}M_{20}^{3} \tag{8}$$

$$\begin{aligned} C_{62} = M_{62} - n6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} \\ +6M_{20}^{2}M_{22} + 24M_{21}^{2}M_{20} \end{aligned} \tag{9}$$

$$\begin{aligned} C_{63} = M_{63} - 9M_{21}M_{42} + 12M_{21}^{3} - 3M_{20}M_{42} \\ -3M_{22}M_{41} + 18M_{20}M_{21}M_{22} \end{aligned} \tag{10}$$

$$X = \begin{bmatrix} |C_{20,norm}| & |C_{21,norm}| & |C_{40,norm}| \\ |C_{41,norm}| & |C_{42,norm}| & |C_{60,norm}| \\ |C_{61,norm}| & |C_{62,norm}| & |C_{63,norm}| \end{bmatrix} \tag{11}$$

where $M_{xy} = \frac{1}{N}\sum_{N} S^{x-y} conj(S)^{y}$ defines the (x+y)-order moment of the modulated symbols S, and conj(S) is the complex conjugate of S. Finally, we normalize all cumulant values by C21 which ensures that the energy of each modulated symbol is equal to unity. Finally, the absolute value of each cumulant feature is appended to create a nine-dimensional feature vector X, as represented in eq. 11, which is passed onwards to a logistic regression classifier.

5. Experimentation and Discussion

We build our DNN models using the Keras framework with the Tensorflow backend [Chollet2015]. The models were trained on an Intel Xeon processor operating at 2.3 GHz. To expedite the training process, we used a Tesla K80 GPU with a 12 GB RAM. All models were optimized using the Adam optimizer with a learning rate of 0.1 with an early stopping callback such that if the model's performance did not improve for 20 epochs, the training process was terminated. The benchmarks results are obtained by first computing cumulant features using the method described in section 5.3 and passing these features to a logistic regression classifier. The classifier is trained with an l2-norm penalty whereas the regularization cost parameter is optimized using grid-search procedure between logarithmically spaced values of 1×10^{-5} and 1×10^{8}.

6.1 Best Performing Model

In Table 1, we report the classification accuracy for the best performing models amongst DNNs and the feature-engineering based model. Given that there are seven modulation schemes to classify between, the chance level accuracy for the dataset is 14.3%, however, all three types of models beat the chance-level baseline. The best performing model achieves an accuracy of 95.03% and is based on the ResNet architecture. The VGG based CNN model follows closely behind with an accuracy of 94.93% whereas the model based on feature engineering achieves only 64.70% only.

Table 1: Classification accuracy for best models from each network architecture type on the validation partition

Model type	***Validation Accuracy***
Cumulant features	64.70 %
CNN	94.93 %
ResNet	95.03 %

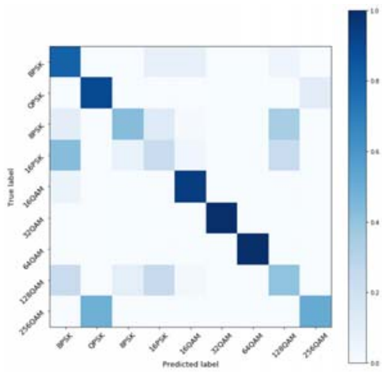
6.2 Model Depth and Classification Accuracy

Depth in deep learning terminology refers to the number of layers for a neural network and is often attributed to the ability of neural networks to learn complex patterns [Sun2016]. We shall now report the influence of model depth on the classification accuracy achieved for each deep learning model for the four network types. To this end, we summarize the classification accuracy achieved by these models between a single and six layers of depth in Table 2. Moreover, we also provide the percent-change in the classification accuracy for every additional layer of depth. Here, one can note a significant increase in classification accuracy between a single layer of depth and two layers of depth, with an improvement of 15.25% for CNN, meanwhile, the ResNet model has a relatively modest improvement of 3.43% in terms of classification accuracy.

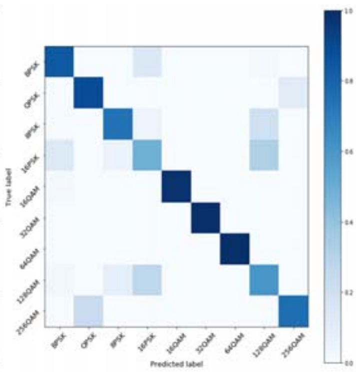
Furthermore, our experiments also suggest that adding more than two layers of depth may not always be useful as we see diminishing returns in terms of classification accuracy after adding more than two layers of depth. This is an important observation because adding more than necessary depth can leave the neural network harder to train and harder to optimize [Du2019]. We believe that this observation is particularly important since it shows that one does not need

Table 2: Classification accuracy with various number of layers of depth

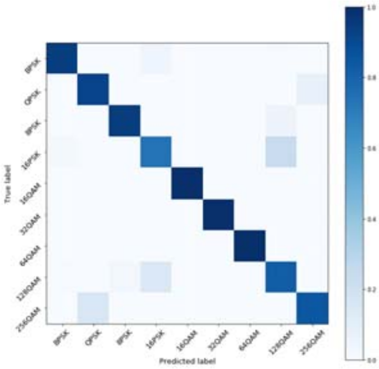
Layers of Depth	*CNN*		*ResNet*	
	Accuracy	*Change*	*Accuracy*	*Change*
1	78.92%	-	90.42%	-
2	93.12%	15.25%	93.63%	3.43%
3	93.68%	0.60%	94.41%	0.83%
4	94.26%	0.62%	94.68%	0.29%
5	94.93%	0.71%	94.88%	0.21%
6	93.39%	-1.54%	95.03%	-0.15%



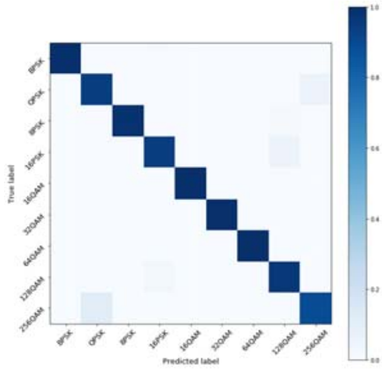
Confusion matrices for 6 layer ResNet for SNR between 0 and 10 dB



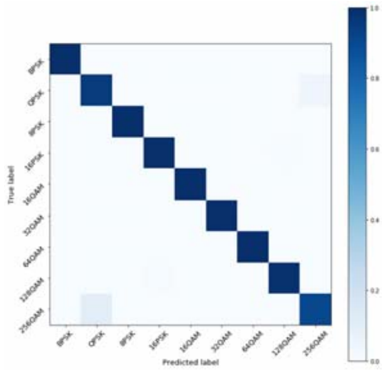
SNR = 2 dB, Overall Accuracy = 89.29%



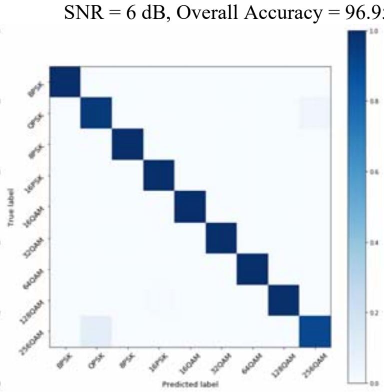
SNR = 4 dB, Overall Accuracy = 91.38%



SNR = 6 dB, Overall Accuracy = 96.95%



SNR = 8 dB, Overall Accuracy = 98.36%



SNR = 10 dB, Overall Accuracy = 98.44%

Fig. 1. Classification accuracy per epoch of best performing models from each of the four network architecture types

deep architectures to achieve good classification accuracy. Simpler architectures, such as

6.3 SNR and Classification Accuracy

In Figure 1, we plot the confusion matrices for ResNet for SNR between 0 and 10 dB (increments of 2 dB). Here one can note that at SNR = 0 dB, a number of modulation schemes are misclassified. For example, 16-PSK is misclassified as BPSK, 256-QAM is misclassified as QPSK, and QPSK is misclassified as 256-QAM. The overall accuracy of the model is also poor at SNR = 0 dB at 69:12%. As the SNR increases, the misclassification rate decreases and therefore, the overall accuracy increases.

It is interesting to note that even at high SNR (as much as 10 dB), the ResNet model continues to misclassify between QPSK and 256-QAM. However, apart from confusing between these two modulation schemes, the classifier maintains high accuracy (as much as 94.44%). These results suggest that deep learning models when deployed for the task of automated recognition of modulation schemes require the received signal at a certain minimum SNR before these models offer high classification accuracy. From our experiments, we determine this SNR to be 8 dB (refer Figure 1-e).

6. Conclusion

In this paper, we investigated the classification performance of four types of DNN architectures i.e., CNN, CNN-FCN, ResNet-C, and ResNet-I for predicting seven types of digital modulation schemes. For our work, we used a publicly available dataset, which enables the reproducibility of our experiments. We reported that, ResNet-I architecture offers the best performance in terms of classification accuracy amongst the four architectures. We also reported that the features learned through a two-layer CNN are linear separable such that one does not need to add multiple layers of fully connected network for the task of classification. This is important because one can reduce the number of training parameters for the DNN and thus the complexity. We observed that while CNN and CNN-FCN can provide a peak classification accuracy relatively close to the ResNet architectures (ResNet-C and ResNet-I), the ResNet architectures offer a more stable learning process. Finally, reported that the classification performance of DNN architectures is limited by the signal-to-noise ratio (SNR). However, when the SNR increases beyond 10 dB, it no longer affects classification accuracy in any significant manner. We believe that this work can provide vital information to other researchers who continue to work towards the development of cognitive radio communication systems.

References

[1] [Cisco2016] Cisco Systems Inc.: *Cisco visual networking index: Forecast and methodology, 2016–2021*. In: Tech. Rep. (2016)

[2] [Cisco2016a] Cisco Systems Inc.: *Cisco visual networking index: Forecast and trends, 2017–2022*, Tech. Rep. (2019)

[3] [Gupta2015] A. Gupta and R. K. Jha: *A survey of 5G network: Architecture and emerging technologies*, IEEE Access, vol. 5, pp. 1206–1232 (2015)

[4] [Ejaz2016] W. Ejaz, A. Anpalagan, M. A. Imran, M. Jo, M. Naeem, S. B. Qaisar, and W. Wang: *Internet of Things (IoT) in 5G wireless communications*, IEEE Access, vol. 4, pp. 10310–10314 (2016)

[5] [Sexton2017] C. Sexton, N. J. Kaminski, J. M. Marquez-Barja, N. Marchetti, and L. A. DaSilva: *5G: Adaptable networks enabled by versatile radio access technologies*, IEEE Communications Surveys and Tutorials, vol. 19, no. 2, pp. 688–720 (2017)

[6] [Mitola1999] J. Mitola and G. Q. Maguire: *Cognitive radio: Making software radios more personal*, IEEE Personal Communications, vol. 6, no. 4, pp. 13–18 (1999)

[7] [Lu2007] C. Lu, F. H. P. Fitzek, and P. C. F. Eggers: *A cooperative scheme enabling spatial reuse in wireless networks*, in Cognitive Wireless Networks, pp. 457–471 (2007)

[8] [Niyato2008] D. Niyato and E. Hossain: *Spectrum trading in cognitive radio networks: A market-equilibrium-based approach*, IEEE Wireless Communications, vol. 15, no. 6, pp. 71–80 (2008).

[9] [Neel2007] J. Neel: *Synthetic symmetry in cognitive radio networks*, in SDR Forum Technical Conference, Denver, CO, USA, pp. 1–6 (2007)

[10] [Rondeau2004] T. W. Rondeau, B. Le, C. J. Rieser, and C. W. Bostian: *Cognitive radios with genetic algorithms: Intelligent control of software defined radios*, In: Software Defined Radio Forum Technical Conference, Phoenix, Arizona, USA, pp. 3–8 (2004)

[11] [Haykin2005] S. Haykin: *Cognitive radio: Brain-empowered wireless communications*, IEEE Journal on Selected Areas in Communications, vol. 23, no. 2, pp. 201–220 (2005)

[12] [Yucek2009] T. Yucek and H. Arslan: *A survey of spectrum sensing algorithms for cognitive radio applications*, IEEE Communications Surveys and Tutorials, vol. 11, no. 1, pp. 116–130 (2009)

[13] [Dobre2007] 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, no. 2, pp. 137–156 (2007)

[14] [Weber2015] C. Weber, M. Peter, and T. Felhauer: *Automatic modulation classification technique for radio monitoring*, Electronics Letters, vol. 51, no. 10, pp. 794–796 (2015)

[15] [Xu2011] 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, no. 4, pp. 455–469 (2011)

[16] [Park2008] C. S. Park, J. H. Choi, S. P. Nah, W. Jang, and D. Y. Kim: *Automatic modulation recognition of digital signals using wavelet features and SVM*, in International Conference on Advanced Communication Technology, Gangwon-Do, South Korea, pp. 1–5 (2008)
[17] [Lv2014] Y. Lv, Y. Liu, F. Liu, J. Gao, K. Liu, and G. Xie: *Automatic modulation recognition of digital signals using CWT based on optimal scales*, in IEEE International Conference on Computer and Information Technology, Xi'an, China, pp. 430–434 (2014)
[18] [LeCun2015] Y. A. LeCun, Y. Bengio, and G. E. Hinton: *Deep learning*, Nature, vol. 521, no. 7553, pp. 436–444 (2015)
[19] [Zhang2016] K. He, X. Zhang, S. Ren, and J. Sun: *Deep residual learning for image recognition*, in IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778 (2016)
[20] [Zhang2017] Y. Zhang, W. Chan, and N. Jaitly: *Very deep convolutional networks for end-to-end speech recognition*, in IEEE International Conference on Acoustics, Speech and Signal Processing, New Orleans, LA, USA, pp. 1–5 (2017)
[21] [Devlin2018] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova: *BERT: Pre-training of deep bidirectional transformers for language understanding*, arXivpreprint:1810.04805v2, vol. 1, no. 1, pp. 1–16 (2018)
[22] [Jagannath2018] J. Jagannath, N. Polosky, D. O'Connor, L. N. Theagarajan, B. Sheaffer, S. Foulke, and P. K. Varshney: *Artificial neural network based automatic modulation classification over a software defined radio testbed*, in IEEE International Conference on Communications, Kansas City, MO, USA, pp. 1–6 (2018)
[23] [OShea2018] T. J. O'Shea, T. Roy, and T. C. Clancy: *Over-the-Air deep learning based radio signal classification,* IEEE Journal on Selected Topics in Signal Processing, vol. 12, no. 1, pp. 168–179 (2018)
[24] [Nandi1998] A. K. Nandi and E. E. Azzouz: *Algorithms for automatic modulation recognition of communication signals*, IEEE Transactions on Communications, vol. 46, no. 4, pp. 431–436 (1998)
[25] [OShea2016] T. J. O'Shea, J. Corgan, and T. C. Clancy: *Convolutional radio modulation recognition networks*, in Engineering Applications of Neural Networks, Aberdeen, UK, 2016, pp. 213–226 (2016)
[26] [Hong2017] D. Hong, Z. Zhang, and X. Xu: *Automatic modulation classification using recurrent neural networks*, in IEEE International Conference on Computer and Communications, Chengdu, China, pp. 695–700 (2017)
[27] [West2017] N. E. West and T. O'Shea: *Deep architectures for modulation recognition*, in IEEE International Symposium on Dynamic Spectrum Access, Baltimore, MD, USA, pp. 1–6 (2017)
[28] [Liu2017] X. Liu, D. Yang, and A. E. Gamal: *Deep Neural Network Architectures for Modulation Classification*, in Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, pp. 915–919 (2017)
[29] [Sainath2015] T. N. Sainath, O. Vinyals, A. Senior, and H. Sak: *Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks*, in IEEE International Conference on Acoustics, Speech and Signal Processing, Brisbane, QLD, Australia, pp. 1–5 (2015)
[30] [Huang2017] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger: *Densely connected convolutional networks*, in IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, pp. 4700–4708 (2017)
[31] [Sabour2017] S. Sabour, N. Frosst, and G. E. Hinton: *Dynamic routing between capsules*, in Advances in Neural Information Processing Systems, Long Beach, CA, USA, pp. 1–11 (2017)
[32] [Zha2019] X. Zha, H. Peng, X. Qin, G. Li, and S. Yang: *A deep learning framework for signal detection and modulation classification*, Sensors, vol. 19, pp. 1–18 (2019)
[33] [Peng2019] S. Peng, H. Jiang, H. Wang, H. Alwageed, Y. Zhou, M. M. Sebdani, and Y. Yao: *Modulation classification based on signal constellation diagrams and deep learning*, IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 3, pp. 718–727 (2019)
[34] [Simonyan2015] K. Simonyan and A. Zisserman: *Very deep convolutional networks for large-scale image recognition*, in International Conference on Learning Representations, San Diego, CA, USA, pp. 1–14 (2015)
[35] [Li2018] X. Li, S. Chen, X. Hu, and J. Yang: *Understanding the disharmony between dropout and batch normalization by variance shift*, arXivpreprint: 1801.05134v1, pp. 1–9 (2018)
[36] [Swami2000] A. Swami and B. M. Sadler: *Hierarchical digital modulation classification using cumulants*, IEEE Transactions on Communications, vol. 48, no. 3, pp. 416–429 (2000)
[37] [Wu2008] H.-C. Wu, M. Saquib, and Z. Yun: *Novel automatic modulation classification using cumulant features for communications via multipath channels*, IEEE Transactions on Wireless Communications, vol. 7, no. 8, pp. 3098–3105, (2008)
[38] [Chang2015] D. C. Chang and P. K. Shih: *Cumulants-based modulation classification technique in multipath fading channels*, IET Communications, vol. 9, no. 9, pp. 828–835 (2015)
[39] [Chollet2015] F. Chollet: *Keras: The python deep learning library* (2015)
[40] [Sun2016] S. Sun, W. Chen, L. Wang, X. Liu, and T. Y. Liu: *On the depth of deep neural networks: A theoretical view*, in AAAI Conference on Artificial Intelligence, Phoenix, Arizona, USA, pp. 1–7 (2016)
[41] [Du2019] S. S. Du, J. D. Lee, H. Li, L. Wang, and X. Zhai: *Gradient descent finds global minima of deep neural networks*, arXiv preprint:1811.03804v4, vol. 1, no. 1, pp. 1–45 (2019)

Muhammad Shehram Shah Syed received his bachelor’s degree in Software Engineering from Mehran University, Pakistan in 2009, and master's in Advanced Computer Science from the University of Strathclyde, UK. He is currently pursuing his doctoral studies at RMIT University

Sajjad Ali Memon received the B.E. degree in telecommunication engineering and the M.E. degree in communication systems networks from Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan, in 2007 and 2011, respectively. He has done his Ph.D. from Dalian University of Technology, Dalian, China, in 2017. He joined Telenor Pakistan as an Operation and Maintenance Engineer, in 2007. He then joined the Department of Telecommunication Engineering, MUET, as a Lab Lecturer, in 2008, became an Assistant Professor in 2011, and became an Associate Professor in 2018. His current research interests include machine learning, Massive MIMO and wireless communication systems.

Erum Saba received her B.E. degree in Electronics Engineering and the M.E. degree in Communication Systems and Networks from Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan, in 2001 and 2010, respectively. She joined Sindh Agriculture University as Lecturer in 2004 and is working in the same department as Assistant Professor since 2014. Her research interests include Signal processing, Control systems, Communication Systems and aritificial intelligence.

Abdul Latif received the B.E. degree in Electrical engineering and the M.E. degree in communication systems networks from Mehran University of Engineering and Technology (MUET), Jamshoro, Pakistan, in 1995 and 2005, respectively. He has done his Ph.D. from Beijing University of Posts Telecom, Beijing, China, in 2013. He joined Pakistan Telecommunication Company Limited as Assistant Divisional Engineer, in 2001. He then joined the Department of Telecommunication Engineering, MUET, as an Assistant Professor, in 2014, became an Associate Professor in 2016. His current research interests include machine learning and Optical Fiber communication systems and Networks).

Aqeel Ahmed has completed his bachelors in Telecommunications Engineering from Mehran University of Engineering and Technology, Jamshoro. He is currently working as a research assistant at the Department of Telecommunication Engineering, Mehran University. His research interests include artificial intelligence, internet of things, and cognitive radio communication systems.