

Sequential Convolutional Recurrent Neural Networks for Fast Automatic Modulation Classification

Kaisheng Liao, Guanhong Tao, Yi Zhong, Yaping Zhang, and Zhenghong Zhang

Abstract—A novel and efficient end-to-end learning model for automatic modulation classification (AMC) is proposed for wireless spectrum monitoring applications, which automatically learns from the time domain in-phase and quadrature (IQ) data without requiring the design of hand-crafted expert features. With the intuition of convolutional layers with pooling serving as front-end feature distillation and dimensionality reduction, sequential convolutional recurrent neural networks (SCRNNs) are developed to take complementary advantage of parallel computing capability of convolutional neural networks (CNNs) and temporal sensitivity of recurrent neural networks (RNNs). Experimental results demonstrate that the proposed architecture delivers overall superior performance in signal to noise ratio (SNR) range above -10 dB, and achieves significantly improved classification accuracy from 80% to 92.1% at high SNRs, while drastically reduces the training and prediction time by approximately 74% and 67%, respectively. Furthermore, a comparative study is performed to investigate the impacts of various SCRNN structure settings on classification performance. A representative SCRNN architecture with the two-layer CNN and subsequent two-layer long short-term memory (LSTM) is developed to suggest the option for fast AMC.

Index Terms—deep learning, convolutional neural networks, recurrent neural networks, automatic modulation classification, spectrum monitoring, cognitive radio.

I. INTRODUCTION

WIRELESS spectrum monitoring over time, space and frequency is important for effective use of the scarce spectral resources in various commercial areas [1]–[5]. As an integral part of wireless spectrum monitoring systems, automatic modulation classification (AMC) is used to recognize modulation types without prior knowledge of the received signals and channel parameters [6]–[8]. AMC has been proven to be an essential capability for transmitter identification, wireless spectrum anomaly detection and radio environment awareness. It improves radio spectrum utilization and opens the possibility of intelligent decision for context-aware autonomous wireless spectrum monitoring systems.

Traditional AMC approaches discussed in literature can be roughly brought down into two main categories: likelihood-based approaches and feature-based approaches [9], [10]. The likelihood-based approaches utilize hypothesis testing theory and form a judgment criterion by analyzing statistical characteristics of signals [11], [12]. In likelihood-based approaches, modulation classification is framed as Bayesian estimation to optimize the probability of classification. However, approaches of this type are not robust in the presence of unknown channel conditions and suffer from heavy computational load

on their practical implementations. Traditional feature-based approaches mainly focus on expert feature extraction and classification criteria [13]–[18]. They utilize expert features such as higher order cyclic moments for modulation classification. It is easy and simple for these approaches to be implemented in practical systems. However, hand-crafting expert features and hard-coding rules for modulation classification make it difficult to scale to new modulation types in non-cooperative scenarios.

Recently, researchers in wireless communications have started to apply deep neural networks to cognitive radio tasks with some success [19]–[32]. The authors in [19], [24] demonstrated that convolutional neural networks (CNNs) trained on time domain in-phase and quadrature (IQ) data significantly outperform conventional expert feature-based approaches. The authors in [20], [23] utilized recurrent neural networks (RNNs) for learning temporal representations to achieve higher classification accuracy than that of the CNNs introduced in [19]. In [21], the authors directly adopted convolutional long short-term deep neural networks (CLDNNs) from voice processing domain. The authors in [33] developed a data-driven fusion method to obtain better classification accuracy using the combination of the two CNNs trained on different datasets. Ramjee et al. [32] performed a comparative study of various typical deep neural networks and reduced the training complexity by reducing the input dimensionality with subsampling techniques.

In autonomous wireless spectrum monitoring systems, on-line learning is fundamental for accommodating new emerging modulation types and complex environmental circumstances. However, those RNN models delivering high classification accuracy suffer from computational complexity and long training time. In this work, we develop a novel and efficient sequential convolutional recurrent neural network (SCRNN) architecture combining parallel computing capability of CNNs with temporal sensitivity of RNNs. Experimental results demonstrate that our approach outperforms the state-of-the-art on classification performance, while significantly improves the rate of convergence compared with the pure CNN and RNN architectures. The code and datasets for all the deep learning models will be made public soon for future research.¹

The rest of the paper is organized as follows. In Section II, an overview of the modulation benchmark dataset is introduced, and the two baseline models are briefly explained. The

¹<https://github.com/kython>

TABLE I
BENCHMARK DATASET PARAMETERS.

Dataset	RadioML2016.10a
Modulations	8 Digital Modulations: BPSK, QPSK, 8PSK, 16QAM, 64QAM, BFSK, CPFSK, and PAM4 3 Analog Modulations: WBFM, AM-SSB, and AM-DSB
Length per sample	128
Signal format	In-phase and quadrature (IQ)
Signal dimension	2×128 per Sample
Duration per sample	$128 \mu s$
Sampling frequency	1 MHz
Samples per symbol	8
SNR Range	$[-20 \text{ dB}, -18 \text{ dB}, -16 \text{ dB}, \dots, 18 \text{ dB}]$
Total number of samples	220000 vectors
Number of training samples	198000 vectors
Number of test samples	22000 vectors

proposed model and the parameters used for training along with other implementation details are clearly stated in Section III. Section IV details the classification results and discusses the advantages of the proposed model. Conclusions and future work are presented in Section V.

II. DATASET AND BASELINES

A. Dataset

In a wireless spectrum monitoring system, the received signal can be typically represented as:

$$r(t) = s(t) * h(t) + n(t) \quad (1)$$

where $s(t)$ denotes the noise free complex baseband envelope of the received signal, and $h(t)$ refers to the time varying impulse response of the transmitted wireless channel. $n(t)$ represents the additive white Gaussian noise (AWGN) reflecting thermal noise. The complex received signal $r(t)$ is commonly sampled in IQ format due to its simplicity.

A typical modulation dataset RadioML2016.10a generated by GNU Radio is used as the benchmark dataset for training and evaluating the performance of the proposed architecture, similar as the MNIST dataset in the vision domain [34]. The dataset follows the signal representation as given in equation 1. Detailed parameter description of the dataset is shown in Table I. Radio channel effects are relatively well characterized in the dataset. Channel imperfections such as multi-path fading, random walk drifting of carrier frequency oscillator and sample time clocks, AWGN, along with unknown scale, translation, and dilation transformation are introduced into the signal in the dataset for reflecting the real electromagnetic environment [34]. The dataset is labeled with both SNR ground truth and modulation types.

B. Baselines

The two models are chosen as the baselines for further comparisons due to their results showing the significant im-

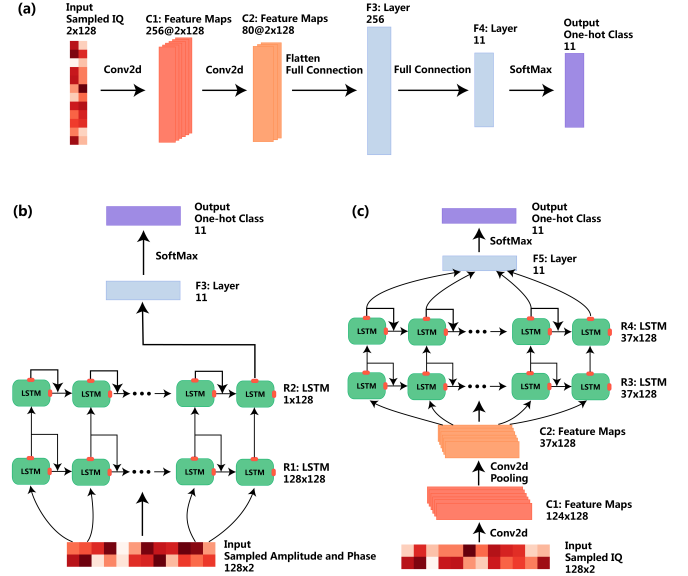


Fig. 1. Schematic diagram of the (a) convolutional neural network (CNN) baseline model, (b) long short-term memory (LSTM) baseline model, and (c) proposed sequential convolutional recurrent neural network (SCRNN) model.

provements upon expert feature-based approaches. Any further improvements should be considered state-of-the-art.

One is the CNN architecture proposed by O'shea et al. [19]. As shown in Fig. 1(a), the baseline model is a 4-layer network made up of two convolutional layers and two dense layers. Each hidden layer utilizes rectified linear unit (ReLU) activation functions and dropout of 50% except for a softmax activation function on the one-hot output layer. The adam optimizer and categorical cross entropy loss function are applied to the base model.

The other baseline model is proposed by Rajendran et al. [23], shown in Fig. 1(b). The model is comprised of two 128-unit long short-term memory (LSTM) layers and an 11-unit dense layer with a softmax activation. The first LSTM layer returns the full sequences while the second one just returns the last state. The dropout is also adopted to reduce overfitting. The adam optimizer and categorical cross entropy loss function are applied to the model. Note that this model learns from the time domain information of the modulation schemes using amplitude-phase format, instead of IQ format.

III. SEQUENTIAL CONVOLUTIONAL RECURRENT NEURAL NETWORKS

A. Motivation

Generally, the received radio signals sampled at discrete time steps are of time domain sequences. In [23], a pure two-layer LSTM architecture is proposed and achieves a good classification accuracy of 86% at high SNRs. However, these models using RNNs suffer from much slower training time than that of the CNNs, due to their computational complexity and unparallel computing capability. Thus, a new novel and efficient SCRNN architecture is proposed with the combination of the speed and lightness of CNNs and the temporal sensitivity of RNNs. Furthermore, as a variant of RNN, LSTM

is adopted instead of simple RNN in the proposed architecture to remember long-term dependencies and avoid the gradient vanishing problem. In SCRNN architectures, the convolutional layers with pooling acting as front-end feature distillation and dimensionality reduction turn the long input sequences into much shorter representations of high-level features, which then become the input for subsequent LSTM layers to learn long-term temporal coherence of modulations.

B. Model Description

Fig. 1(c) provides the illustration of the proposed SCRNN architecture. As schematically shown in Fig. 1(c), the first and second convolutional layers each contain 128 5-tap filters except for the first one followed by a max-pooling layer with a pooling size of 3. The layer 3 and layer 4 are LSTM layers composed of 128 units each, and both return the full sequences. The last dense layer contains 11-class neurons representing the modulation schemes.

ReLU activation functions are applied to the convolutional and LSTM layers. The last dense layer utilizes a softmax activation to achieve modulation classification. Dropout regularization combined with max norm has been proven to be of better performance for preventing overfitting. The adam optimizer with a learning rate of 0.001 and categorical cross entropy loss function are adopted.

C. Implementation details

The total 220000 samples in the RadioML2016.10a dataset are split into two, one training set of 198000 (90%) samples and the other test set of 22000 (10%) samples. The dataset is split equally among all considered modulation types using the stratified sampling strategy. Instead of extracting the amplitude and phase features of the signals manually in advance [23], we adopted IQ components as input directly. A batch size of 128 is used on each training epoch and the early stop strategy is adopted.

All training and prediction are implemented in Keras library [35] on the backend of TensorFlow [36]. The Nvidia Cuda enabled Tesla K80 is used to speed up the calculation.

IV. RESULTS AND DISCUSSION

The classification performance of the models on the benchmark dataset is discussed in this section. We inspect and compare the classification accuracy and rate of convergence between the baseline models and the proposed SCRNN model. In addition, the varying kernel sizes, kernel types and layer depths are further investigated to find the optimal SCRNN architecture.

The classification accuracy of all the models are presented in Fig. 2. It can be seen that the proposed SCRNN model delivers a significantly improved accuracy of 92.1% at high SNRs. The CNN and LSTM model as baselines are compared to the proposed SCRNN model. It shows that the SCRNN model consistently achieves higher accuracy than the other two baselines in the SNR range from -10 dB to 18 dB, and significantly outperforms the CNN baseline model by 12%

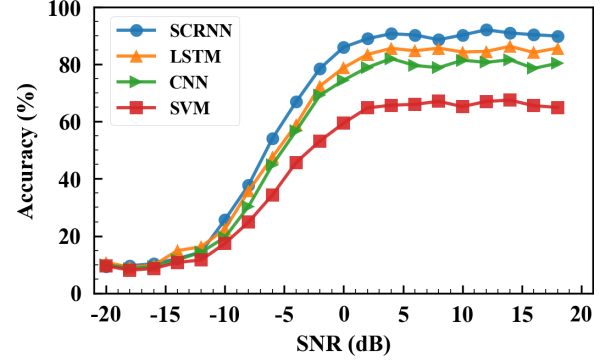


Fig. 2. Classification accuracy comparison of the proposed SCRNN model with others on the benchmark dataset.

and the LSTM baseline model by 6% improvement at high SNRs. Additionally, it is observed that the proposed SCRNN model achieves exceeding performance than that of the CNN and LSTM baseline models in the SNR range from -10 dB to 0 dB, where the two baseline models behave nearly the same. It implies that the convolutional layers of the SCRNNs serving as feature distillation boost the learning ability of the temporal features under low SNR circumstances. The traditional support vector machine (SVM) approach showing poor classification performance is also summarized in Fig. 2 for comparison. All models are fed with the same training and test data of IQ format for this comparison except for the LSTM model with amplitude-phase format.

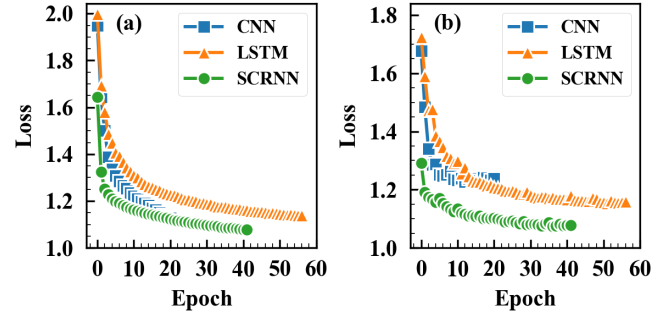


Fig. 3. Training history including the (a) training loss and (b) validation loss between the baseline models and the proposed SCRNN model.

Fig. 3 shows the training history including the (a) training loss and (b) validation loss compared between the baseline models and the proposed SCRNN model. According to the training history, the LSTM baseline model achieves the second less loss value but remains the lowest rate of convergence; the CNN baseline model obtains faster rate of convergence but yields the largest loss value, while the proposed SCRNN model retains the fastest rate of convergence and achieves the least loss value among the three. Moreover, the training and prediction time of the proposed 2-layer SCRNN model are drastically reduced to only 280 seconds per epoch and $661 \mu s$ per sample respectively, compared to 800 seconds per epoch and $2000 \mu s$ per sample of the 2-layer LSTM baseline model, as shown in Table II. These are fairly consistent with the insight that the convolutional layers with pooling before

TABLE II
TRAINING AND PREDICTION TIME COMPARISON BETWEEN THE TWO
BASELINE MODELS AND THE THREE SCRNN MODELS.

Models	CNN	LSTM	1-Layer SCRNN	2-Layer SCRNN	3-Layer SCRNN
Training Time per Epoch (s)	30	800	800	280	90
Training Epochs	20	56	57	41	57
Total Training Time (s)	600	44800	45600	11480	5130
Prediction Time (μ s/sample)	1000	2000	641	661	751

RNN serve as feature distillation and dimensionality reduction, analogous to front-end matched filters, synchronizer and sampler for temporal features in typical wireless systems. To gain intuition on what convolution layers are learning in SCRNN architectures, the response patterns of the 128 filters learned by the first convolutional layer are illustrated in Fig. 4, showing that some filters encode expert-like patterns (i.e. BPSK-like pattern in row 1 column 6) and others even encode more complicated patterns. It further confirms that the convolutional layers of the SCRNNs act as front-end feature distillation with coherent features refined and redundant features filtered out, enabling the improved rate of convergence.

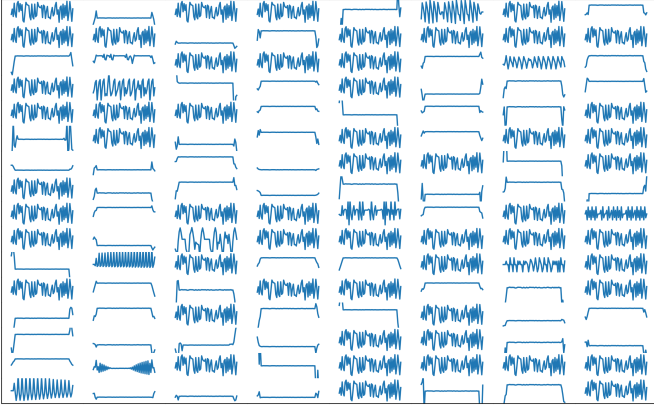


Fig. 4. Response patterns of the 128 filters learned by the first convolutional layer of the SCRNN.

To gain more insight into the SCRNN architecture, we further investigate the effects of various SCRNN structure settings varying CNN kernel sizes, CNN layer depths, CNN kernel numbers, RNN types and RNN layer depths on classification performance.

As shown in Fig. 5(a), varying the CNN kernel sizes of the SCRNNs has minimal impact on classification performance. The architecture with kernel size of 5 produces slightly better classification accuracy than others in SNR range from 0 dB to 18 dB, while the architecture with kernel size of 3 leads to marginally higher classification accuracy in SNR range from -10 dB to -6 dB. The kernel size of 5 is used for the remaining experiments.

Fig. 5(b) proves that the reduction of the input dimensionality for subsequent LSTM layers in the SCRNN architectures

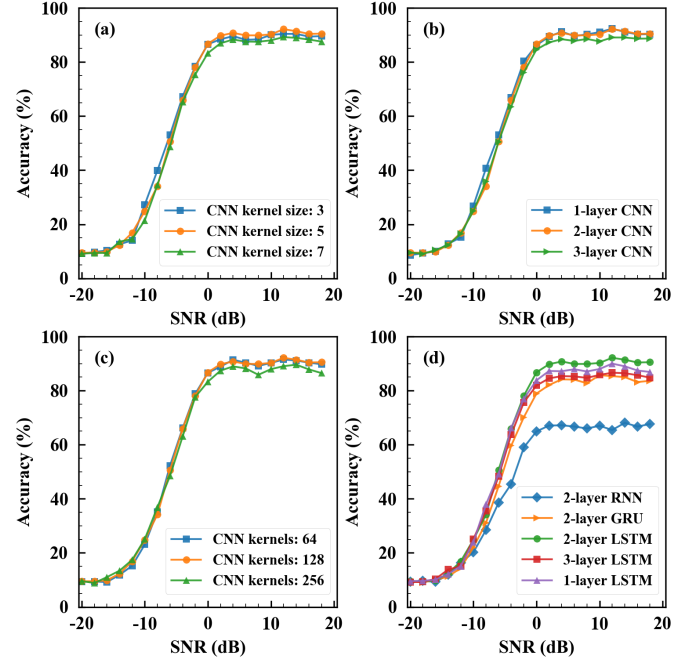


Fig. 5. Classification performance of different SCRNN structure configurations for varying (a) CNN kernel sizes, (b) CNN layer depths, (c) CNN kernel numbers, (d) RNN types and layer depths.

shows very limited effects on classification performance. As shown in Fig. 1 and Table II, the 2-layer SCRNN model reduces the dimensionality by a factor of 3. This leads the training and prediction time reduced by 74% and 67% respectively, while the performance remains nearly the same. However, the performance of the LSTM baseline model starts to decay significantly when reducing the input dimensionality [32]. It is implied that the SCRNN architecture is much more robust to dimensionality reduction and training and prediction time minimization. Thus, it makes possible for deploying online learning model on autonomous wireless spectrum monitoring systems.

Fig. 5(c) provides that the 64-kernel and 128-kernel structures deliver the very similar performance, while the performance of 256-kernel structure starts to drop due to the overfitting. Fig. 5(d) shows the different settings of RNN types and layer depths in the SCRNN architectures. It can be observed that the performance of the LSTM type is apparently superior to that of the gated recurrent unit (GRU) and simple RNN type. Experimental results of varying LSTM layer depths suggest that the 2-layer LSTM of the SCRNN achieves the best classification accuracy. Therefore, the optimal SCRNN architecture with the 2-layer CNN and subsequent 2-layer LSTM is recommended for online learning.

To evaluate how classification performance varies with SNRs, confusion matrices of the optimal SCRNN model at various SNRs are investigated. For a confusion matrix, each column represents the predicted modulation type and each row represents the real modulation type. The numerical value on each grid denotes the prediction probability of the corresponding modulation type.

As illustrated in Fig. 6, the diagonals become gradually

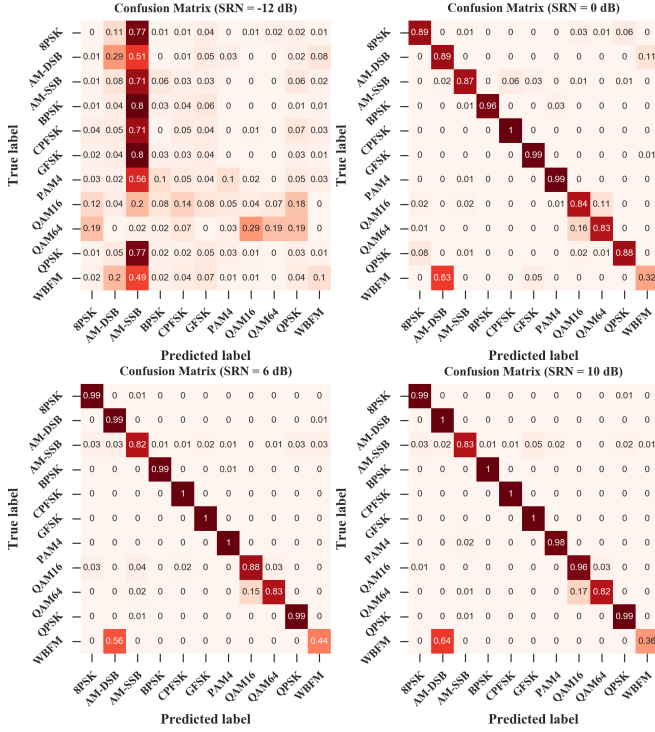


Fig. 6. Confusion matrices for the optimized SCRNN architecture on the benchmark dataset at various SNRs.

sharper with increasing SNR, yet two primary confusions exist even at high SNRs. One is among the analog modulations. This is mainly due to the silent period exiting in the analog audio signal [19]. The other is between QAM16 and QAM64 as the former is a subset of the latter.

V. CONCLUSION

In this paper, a novel and efficient SCRNN architecture for AMC has been developed. Compared with the pure CNN and LSTM baseline models, the proposed architecture takes full advantage of the complementarity of CNNs and RNNs. Thus, it makes the classification accuracy deliver the state-of-the-art performance, improved from 80% to 92.1% at high SNRs. Furthermore, the training and prediction time of the proposed architecture are significantly reduced by approximately 74% and 67% respectively with negligible loss in classification accuracy, paving the way for deployment of online learning models on autonomous wireless spectrum monitoring systems. Additionally, a comparative study of various structure settings of SCRNNs has been performed and a representative SCRNN architecture with the 2-layer CNN and the subsequent 2-layer LSTM is recommended for fast AMC. Future work will focus on validation on radio signals with varying symbol rates and bandwidths. Second, unsupervised or deep reinforcement learning approaches for AMC should be investigated due to the lack of necessary signal labels in real wireless spectrum monitoring systems. Finally, stream learning without requiring to retrain the entire network from scratch is also a worthy direction for future research.

REFERENCES

- [1] M. Höyhty, A. Mämmelä, M. Eskola, M. Matinmikko, J. Kalliovaara, J. Ojanen, J. Suutala, R. Ekman, R. Bacchus, and D. Roberson, "Spectrum occupancy measurements: A survey and use of interference maps," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2386–2414, 2016.
- [2] S. Barber, R. Petruschka, and E. de Castro, "Using wireless network access points for monitoring radio spectrum traffic and interference," Mar. 18 2004. US Patent App. 10/430,731.
- [3] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE communications surveys & tutorials*, vol. 11, no. 1, pp. 116–130, 2009.
- [4] S. Zheng, S. Chen, L. Yang, J. Zhu, Z. Luo, J. Hu, and X. Yang, "Big data processing architecture for radio signals empowered by deep learning: Concept, experiment, applications and challenges," *IEEE Access*, vol. 6, pp. 55907–55922, 2018.
- [5] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE Journal on selected areas in communications*, vol. 31, no. 11, pp. 2209–2221, 2013.
- [6] C. Weber, M. Peter, and T. Felhauer, "Automatic modulation classification technique for radio monitoring," *Electronics Letters*, vol. 51, no. 10, pp. 794–796, 2015.
- [7] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 47–52, 2007.
- [8] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and knn," *IEEE Transactions on wireless communications*, vol. 11, no. 8, pp. 2742–2750, 2012.
- [9] 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.
- [10] 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.
- [11] 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, 2010.
- [12] W. Wei and J. M. Mendel, "Maximum-likelihood classification for digital amplitude-phase modulations," *IEEE transactions on Communications*, vol. 48, no. 2, pp. 189–193, 2000.
- [13] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," *IEEE Circuits and Systems Magazine*, vol. 9, no. 2, pp. 27–45, 2009.
- [14] 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.
- [15] 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 *2008 10th International Conference on Advanced Communication Technology*, vol. 1, pp. 387–390, IEEE, 2008.
- [16] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Transactions on communications*, vol. 48, no. 3, pp. 416–429, 2000.
- [17] S.-Z. Hsue and S. S. Soliman, "Automatic modulation classification using zero crossing," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 137, pp. 459–464, IET, 1990.
- [18] S. S. Soliman and S.-Z. Hsue, "Signal classification using statistical moments," *IEEE Transactions on Communications*, vol. 40, no. 5, pp. 908–916, 1992.
- [19] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*, pp. 213–226, Springer, 2016.
- [20] D. Hong, Z. Zhang, and X. Xu, "Automatic modulation classification using recurrent neural networks," in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 695–700, IEEE, 2017.
- [21] N. E. West and T. O'Shea, "Deep architectures for modulation recognition," in *2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, pp. 1–6, IEEE, 2017.
- [22] A. Ali, F. Yangyu, and S. Liu, "Automatic modulation classification of digital modulation signals with stacked autoencoders," *Digital Signal Processing*, vol. 71, pp. 108–116, 2017.

- [23] S. Rajendran, W. Meert, D. Giusiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 3, pp. 433–445, 2018.
- [24] 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.
- [25] M. Sadeghi and E. G. Larsson, "Adversarial attacks on deep-learning based radio signal classification," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 213–216, 2018.
- [26] D. Zhang, W. Ding, B. Zhang, C. Xie, H. Li, C. Liu, and J. Han, "Automatic modulation classification based on deep learning for unmanned aerial vehicles," *Sensors*, vol. 18, no. 3, p. 924, 2018.
- [27] C.-F. Teng, C.-C. Liao, C.-H. Chen, and A.-Y. A. Wu, "Polar feature based deep architectures for automatic modulation classification considering channel fading," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 554–558, IEEE, 2018.
- [28] J. Sun, G. Wang, Z. Lin, S. G. Razul, and X. Lai, "Automatic modulation classification of cochannel signals using deep learning," in *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)*, pp. 1–5, IEEE, 2018.
- [29] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018.
- [30] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
- [31] S. Zheng, P. Qi, S. Chen, and X. Yang, "Fusion methods for cnn-based automatic modulation classification," *IEEE Access*, 2019.
- [32] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. Gamal, and Y. C. Eldar, "Fast deep learning for automatic modulation classification," *arXiv preprint arXiv:1901.05850*, 2019.
- [33] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 4074–4077, 2019.
- [34] T. J. O'shea and N. West, "Radio machine learning dataset generation with gnu radio," in *Proceedings of the GNU Radio Conference*, vol. 1, 2016.
- [35] F. Chollet *et al.*, "Keras: The python deep learning library," *Astrophysics Source Code Library*, 2018.
- [36] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pp. 265–283, 2016.