Automatic Modulation Classification: A Deep Architecture Survey

THIEN HUYNH-THE1, (Member, IEEE), QUOC-VIET PHAM2, (Member, IEEE), TOAN-VAN NGUYEN3, (Member, IEEE), THANH THI NGUYEN4, RUKHSANA RUBY5, (Member, IEEE), MING ZENG6, (Member, IEEE), AND DONG-SEONG KIM1, (Senior Member, IEEE)

1Department of IT Convergence, Kumoh National Institute of Technology, Gumi, Gyeongsangbuk-do 39177, Republic of Korea
2Korean Southeast Center for the 4th Industrial Revolution Leader Education, Pusan National University, Busan 46241, Republic of Korea
3Department of Electrical and Computer Engineering, Utah State University, Logan, UT 84321, USA
4School of Information Technology, Deakin University (Burwood Campus), Burwood, VIC 3125, Australia
5College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
6Department of Electrical and Computer Engineering, Laval University, Quebec, QC G1V 0A6, Canada

Corresponding authors: Thien Huynh-The (thienht@kumoh.ac.kr) and Dong-Seong Kim (dskim@kumoh.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) through Creativity Challenge Research-Based Project under Grant 2019R1I1A1A01063781, and in part by the National Research Foundation of Korea (NRF), Institute for Information and Communications Technology Planning and Evaluation (IITP) by the Ministry of Education, Science and Technology under Grant 2018R1A6A1A03024003 and Grant IITP-2020-2020-0-01612.

ABSTRACT Automatic modulation classification (AMC), which aims to blindly identify the modulation type of an incoming signal at the receiver in wireless communication systems, is a fundamental signal processing technique in the physical layer to improve the spectrum utilization efficiency. Motivated by deep learning (DL) high-impact success in many informatics domains, including radio signal processing for communications, numerous recent AMC methods exploiting deep networks have been proposed to overcome the existing drawbacks of traditional approaches. DL is capable of learning the underlying characteristics of radio signals effectively for modulation pattern recognition, which in turn improves the modulation classification performance under the presence of channel impairments. In this work, we first provide the fundamental concepts of various architectures, such as neural networks, recurrent neural networks, long short-term memory, and convolutional neural networks as the necessary background. We then convey a comprehensive study of DL for AMC in wireless communications, where technical analysis is deliberated in the perspective of state-of-the-art deep architectures. Remarkably, several sophisticated structures and advanced designs of convolutional neural networks are investigated for different data types of sequential radio signals, spectrum images, and constellation images to deal with various channel impairments. Finally, we discuss some primary research challenges and potential future directions in the area of DL for modulation classification.

INDEX TERMS Automatic modulation classification, convolutional neural networks, deep learning architectures, deep neural networks, recurrent neural networks, wireless communications.
modulation format is selected depending on system specifications and channel conditions. Automatically identifying the modulation types of received signals allows the receiver to demodulate the signal, and thus the development of an efficient algorithm for modulation identification is the priority in many software-defined radio-based communications. Automatic modulation classification (AMC), a preceding process of signal demodulation in the physical layer [5], is currently attracting more attention from signal processing and communication societies. Fundamentally, AMC aims to classify the modulation type of an incoming signal at the receiver, which typically plays as a multi-class decision-making task from the perspective of artificial intelligence (AI). Concretely, the underlying radio characteristics, including the information of modulation type, can be obtained by conventional feature engineering techniques (e.g., feature extraction and feature selection) to learn a classification model using supervised or unsupervised learning [6]. Nevertheless, AMC has to confront many challenging issues [7] such as increasing number of modulation formats, intra-class discrimination of higher-order digital modulations, the strong channel impairments.

In the last decades, numerous advanced analog and digital modulation techniques have been employed in communication systems to achieve a good balance between spectrum efficiency and transmission reliability [8]. In analog communication systems, a transmission signal is encoded by using analog modulations, such as amplitude modulation (AM), phase modulation (PM), and frequency modulation (FM). Typically, an analog modulation technique encodes an analog baseband signal (so-called the source signal) onto a high-frequency periodic waveform (so-called the carrier signal). Compared with analog modulations, digital modulations are more preferable in terms of usage thanks to better coordination with digital data and stronger robustness against interference. In the communication systems with digital modulations as given in Fig. 3, the source signal is first digitized by sampling and quantization, and the resulted digital signal is then coded to improve data security and reduce transmission errors before passing to a digital modulator. Several commonly used digital modulations are amplitude-shift keying (ASK), phase-shift keying (PSK), frequency-shift keying (FSK), pulse amplitude modulation (PAM), amplitude and phase-shift keying (APSK), and quadrature amplitude modulation (QAM). In the modulation process, different waveform characteristics of the carrier signal (such as amplitude, frequency, phase, and a combination of amplitude and phase) can be modified based on the pre-defined modulation technique. Over propagation channels, the most appropriate modulation of an incoming signal is determined by inferring its radio characteristics with a learned AI model.

A. STATE-OF-THE-ART APPROACHES

Numerous AMC methods have been proposed to assist dynamic spectrum access and intelligent spectrum management. This section briefly reviews the existing traditional AMC methods, where most of them are basically categorized into the likelihood-based (LB) and feature-based (FB) approaches with the general processing flow illustrated in Fig. 2.

1) SIGNAL MODEL

We first present a regular signal model in single-input single-output single-carrier systems with the presence of channel deterioration. Let \( x[n, H_k] \) be the noiseless signal under transmission channel impairments that can be written as

\[
x[n, H_k] = A_n e^{i(2\pi f_o n T + \theta_n)} \sum_{k=-\infty}^{\infty} x[k] h(n T - k T + \epsilon_T),
\]

where \( A_n \) is the signal amplitude of symbol \( n, f_o \) is the carrier frequency offset, \( \theta_n \) is the varying phase offset, \( T \) is the symbol spacing (or interval), \( h(\cdot) \) denotes the synthetic effect of the residual baseband channel, \( x[k] \) denotes the symbol sequence of the original data over a specific modulation scheme, and \( \epsilon_T \) is the timing offset between the transmitter and the receiver. Subsequently, the complex envelope of the received radio signal \( y[n] \) is expressed as follows:

\[
y[n] = x[n, H_k] + w[n],
\]

where \( w[n] \) is the additive white Gaussian noise (AWGN). In communication systems, an AMC approach aims to predict the modulation format of \( x[n, H_k] \) precisely without the channel state information \( H_k \) by learning the underlying features of the received signal \( y[n] \).

2) TRADITIONAL APPROACHES

Many AMC methods in this group have deployed probabilistic frameworks in likelihood-based approaches and traditional machine learning frameworks in feature-based approaches. Generally, the likelihood-based approaches apply probability theories and hypothesis models to solve
modulation identification problems under the conditions of known and unknown channel information [9]. Although the likelihood-based approaches can reach the optimal classification accuracy with the perfect knowledge of signal model and channel model (which cannot be obtained in the real world), they require high computation complexity to estimate model parameters [10], [11]. By following a regular machine learning (ML) framework for classification task, the feature-based approaches are more favorable to deploy in practical systems compared to the likelihood-based approaches, thanks to their relatively easy implementation and low complexity [12]. Despite being flexible with different channel models, the feature-based approaches face some major drawbacks: weakly discriminative experience of handcrafted features and limited learning capacity of traditional classification algorithms [13]–[15].

3) REVOLUTIONARY APPROACHES
A few years ago, inspired by the great explosion of DL with unprecedented success in different fields, from computer vision [16], [17] to communications [18], [19], several revolutionary approaches have exploited different DL architectures, such as deep neural networks (DNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and convolutional neural networks (CNNs or ConvNets) to improve the performance of modulation classification significantly. Compared with conventional machine learning, DL has presented the essential advantages of automatic feature extraction and high learning capacity [20], [21], thus increasing the classification accuracy of higher-order modulation formats under a synthetic channel deterioration [22]. Besides that, effectively processing big data enables DL to be deployed for modulation classification in Internet-of-Things (IoT) systems [23], where a vast amount of data is generated by edge devices.

With regards to AMC, few surveys have been conducted over the last decades. In [7], the intensive survey of likelihood-based and feature-based approaches has been provided along with the summary of numerical results and the discussion about research trends. As being more contemporary compared to [7], the work [24] first described the fundamental of deep networks and then briefly reviewed DL-based AMC approaches. However, in [24], insightful analysis and discussion about deep architectures were ignored. Consequently, limitations and advantages of deep networks for specific channel conditions were presented insufficiently. Some broader surveys of DL for several challenging tasks in wireless communications have been accomplished recently. Mao et al. [25] reviewed the use of DL for multiple tasks in the physical layer, such as interference alignment, jamming resistance, physical coding, and modulation classification. Li et al. [26] delivered a short survey of DL-enabled wireless signal identification and modulation recognition for intelligent spectrum monitoring and management in fifth-generation (5G) and IoT networks. The review conducted by Jid et al. [27] mainly focused on the application of DL for AMC in single-antenna and multi-antenna systems from the communication perspective. A recent survey [28] covered four fundamental topics of intelligent radio signal processing, including modulation classification, signal detection, beamforming, and channel estimation. Although the above-mentioned surveys concluded that DL can improve the performance of modulation classification to obtain high reliability in communication systems, these failed to analyze deep architectures in a comprehensive manner and point out the way how deep networks with sophisticated-designed architecture can boost the accuracy while keeping an acceptable complexity.

B. CONTRIBUTIONS
In this work, we first review traditional AMC methods, including the likelihood-based approaches and feature-based approaches, wherein conventional feature extraction algorithms and classification algorithms are applied to learn modulation patterns. Under the umbrella of DL techniques, the fundamental concepts of different deep architectures are provided systematically from basic layers to advanced processing units. Then, we review state-of-the-art AMC methods, wherein DL is exploited as the core technology to improve the overall performance of modulation classification. Finally, we remark on several challenging issues and future research directions on the topic of AMC. In a nutshell, the primary contributions of this paper are summarized as follows.

• We briefly review traditional AMC methods, where the underlying idea of likelihood-based and feature-based approaches are presented. Accordingly, their inherent drawbacks are pointed out for discussion.
• We present the fundamental concepts of FFNN, RNN, LSTM, and CNN architectures. Remarkably, the operating principles of different layers and processing units are described in a comprehensive manner.
• We survey revolutionary AMC approaches, where various deep models are developed as classifiers to not only overcome the limitations of traditional AMC methods but also improve the performance in terms of accuracy and complexity. Notably, several sophisticated-designed networks with attached architectural diagrams are analyzed and examined profoundly for knowledge enlightenment.
• We reveal a number of practical challenges and recommend potential directions to improve the efficiency of AMC in terms of accuracy and complexity.

This paper is organized as follows. In Section II, we present the fundamental concepts of DL architectures. Next, Section III discusses state-of-the-art AMC methods which have exploited various DL architectures (e.g., FFNNs, RNNs, LSTM, and CNNs) for different data types (such as sequential radio signals, spectrum images, and constellation images) to enhance learning efficiency and improve classification accuracy. Challenges and future directions of AMC are drawn in Section IV, and finally we conclude the paper in
Section V. The schematic outline of this survey is shown in Fig. 4.

II. DEEP LEARNING: FUNDAMENTAL CONCEPTS

A. FEEDFORWARD NEURAL NETWORK

Artificial neural network (ANN), a very broad term that usually represents any form of DL model, is specified by a group of multiple neurons in each layer. Since the network inputs are processed only in the forward direction without going backward, ANN is also known as FFNN. ANNs can be either shallow with only one hidden layer (i.e., one layer between the input and the output) or deep with more than one layer (i.e., multiple hidden layers). A deep network with multiple hidden layers is generally so-called FFNN which allows passing information through numerous intermediate nodes, from the input nodes to the output nodes [29], without going backward and the links between layers are one way in the forward direction. The architecture of a simple FFNN with one input layer, two hidden layers, and one output layer is shown in Fig. 5. Despite representing a wide range of stacked neural networks, FFNNs for existing AMC methods typically involve more than two hidden layers with different numbers of connected nodes. The nonlinearity in FFNNs is represented by activation functions, which help the network to learn any complex relationship between the input and the output. Some activation functions commonly used in FFNNs include logistic sigmoid function

$$ f(x) = \frac{1}{1 + e^{-x}}, \quad (3) $$

and the tangent hyperbolic (tanh) function

$$ f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad (4) $$

which results output in the range $[0 \ 1]$ and $[-1 \ 1]$, respectively.

Despite being more interpretable, FFNNs pose several limitations that can reduce the overall network performance [30]. At first, when processing high-dimensional unstructured data like image, it should convert to one-dimensional vector. That leads the drastically increasing number of trainable parameters along with the image size (or resolution) and the loss
of spatial features during the model learning stage. Secondly, when more layers are designed in a very deep NN, the gradient of network decreases exponentially to cause ineffective network training and possibly lead to overall inaccuracy of the whole network [31].

B. RECURRENT NEURAL NETWORK

Different from a typical multilayer network with feed-forward connections, RNN is extended with the concept of recurrent connections (so-called recurrent edges) to feed information back into prior layers (or into the same layer). At each time-step of passing input through a basic RNN with a chain-like structure as shown in Fig. 6, nodes (so-called memory cells) process the activation results of the current input vector and of the previous-state hidden nodes. This mechanism allows RNNs to maintain past information for processing the current input and to capture the long-range time dependencies of the input data [32], whereby favorably handling many challenging problems related to time series data, text data, and audio data [33].

Besides learning the sequential information from the input data, another advantage of RNNs is parameter sharing (i.e., shares the parameters across difference time-steps in the architecture) which reduces the network size and computational cost. As shown in Fig. 6, three weight matrices $W_e, W_h, W_o$ are shared across all the time-steps. RNNs adversely process very long sequences if configuring the activation functions tanh or rectified linear unit (ReLU) as shown in Fig. 7a, the components at time-step $t$ can be written with the input $x_t$ and the hidden state $h_{t−1}$ as follows.

$$f_t = σ (W_f x_t + R_f h_{t−1} + b_f),$$

forget gate: $f_t = σ (W_f x_t + R_f h_{t−1} + b_f),$

Input gate: $i_t = σ (W_i x_t + R_i h_{t−1} + b_i),$

Cell candidate: $g_t = \tanh (W_c x_t + R_c h_{t−1} + b_c),$

Output gate: $o_t = σ (W_o x_t + R_o h_{t−1} + b_o),$ \hspace{1cm} (6)

where $σ$ denotes the signal activation function; the matrices $W, R, \text{and} \ b$ are the sets of the input weights, the recurrent weights, and the biases of each component, respectively. The cell state at time-step $t$ is defined by

$$c_t = f_t * c_{t−1} + i_t * g_t.$$ \hspace{1cm} (7)

where $*$ denotes the Hadamard product (i.e., element-wise product). The hidden state $h_t$ is then updated by

$$h_t = o_t * \tanh (c_t).$$ \hspace{1cm} (8)

Another dramatic variation of recurrent unit is gated recurrent unit (GRU), in which the forget and input gates are merged into a single component [37], called update gate. Some other changes on the cell state and hidden state are given to simplify the unit’s structure. From the structure of a basic GRU shown in Fig. 7b, the components are calculated as

Update gate: $z_t = σ (W_z x_t + R_z h_{t−1} + b_z),$

Reset gate: $r_t = σ (W_r x_t + R_r h_{t−1} + b_r),$

Cell candidate: $\tilde{h}_t = \tanh (W_h x_t + R_h (r_t * h_{t−1}) + b_h),$

Output: $h_t = (1−z_t) * h_{t−1} + z_t * \tilde{h}_t.$ \hspace{1cm} (9)

Compared with LSTM, GRU owns lower-cost computation and easier implementation [38].

C. CONVOLUTIONAL NEURAL NETWORK

One of the most popular and successful deep architectures is CNN, which employs convolution operations to learn higher-order features in the data [39]–[41]. With the capability of processing high-dimensional unstructured data, CNNs are specifically suitable for images as inputs [42], although they are further exploited for other applications with text, signals, and other continuous responses [43]. A basic CNN consists of an input layer, multiple hidden layers, and an output layer, in which a hidden layer typically involves a convolutional layer followed by an activation layer and other additional layers, such as pooling layer, fully connected layer (so-called dense layer), and normalization layer. Compared with other deep network architectures which cannot share any connections and produce outcomes, CNNs are more preeminent in learning meaningful features from raw data [3], [44].

As the core concept of CNN, the convolution operation (a.k.a., feature detector) calculates higher-level features from raw input data or lower-level features with convolution kernels (so-called filters). Typically, convolution is the dot product of the weights (hyper-parameters in CNNs) of a given kernel and the elements of an input map within a receptive
field identified by a spatial size (height \times width), where the depth size of kernel is regularly identical to the number channels of input map. The mathematical formulation of computing convolution at a coordinate \((i,j)\) is as follows:

\[
\text{conv} (W, I)_{i,j} = \sum_{m,n,k} W_{m,n,k} I_{i+m,j+n,k} + b,
\]

where \(W \in \mathbb{R}^{M \times N \times K}\) and \(I\) denote the kernel and the input, respectively, and \(b\) is the scalar bias. Besides ReLU, some other activation functions usually considered in designing CNN architectures are leaky ReLU and exponential linear unit (eLU) as follows:

\[
f(x) = \text{leakyReLU}(x) = \begin{cases} 
x & \text{if } x \geq 0, \\
\beta x & \text{if } x < 0,
\end{cases}
\]

(11)

where a multiplication with a fixed scalar \(\beta\) is given for any negative input,

\[
f(x) = \text{eLU}(x) = \begin{cases} 
x & \text{if } x \geq 0, \\
\alpha (e^x - 1) & \text{if } x < 0,
\end{cases}
\]

(12)

where \(\alpha\) is the nonlinear parameter and the function performs an exponential nonlinearity on any negative input. Compared with ReLU, leaky ReLU and eLU handle more effectively the zero dying problem (i.e., the learning is stopped with negative neurons because of zero gradient). In some applications, eLU promingly drives the network convergence more quickly with a more accurate result compared with leaky ReLU.

Pooling layers are commonly arranged between successive convolutional layers to reduce the spatial size of feature maps. By dividing the input into rectangular pooling regions (defined by pool size), a max pooling layer performs downsampling via resulting the maximum of each region while an average pooling calculates the mean value of each region. Despite reducing the number of parameters to be learned in the following convolutional layers, pooling layers also mitigate overfitting issue. In several classification networks [2], a global average pooling layer performs down-sampling by computing the height and width dimensions along the channel dimension of the input, which can be used before the final fully connected layer to significantly reduce the network size without sacrificing performance.

### References

In numerous typical CNNs [45]–[47], connecting all neurons in a fully connected layer to all neurons in the previous layer allows the network to summarize the locally informative features learned in many former layers to classify the input. The number of neurons in the last fully connected layer is identical to the number of classes in a given dataset. For general multi-class classification tasks, a softmax layer, which usually follows the last fully connected layer, applies a softmax function to calculate a probability for every possible class from the activation outcomes of the previous layer. The class with the maximum probability implies the classification output (i.e., the output of the entire network). In the training process, CNNs usually calculate the cross entropy loss for multi-class classification as

\[
\mathcal{L} = - \sum_{i=1}^{N_S} \sum_{j=1}^{n_c} u_{ij} \ln (v_{ij}),
\]

(13)

where \(N_S\) is the number of training samples, \(u_{ij}\) denotes the ground-truth of the \(i\)-th sample associated with the \(j\)-th class, and \(v_{ij}\) denotes the output inferred by the network for the class \(j\) of the sample \(i\).

Besides the aforementioned layers, some additional layers can be employed to improve the learning efficiency of CNNs, such as batch normalization (bn) layer, element-wise addition layer, and concatenation layer (including spatial-wise and depth-wise concatenation). Batch normalization is usually utilized in CNNs to accelerate the network training convergence and reduce the sensitivity to network initialization by reducing the internal co-variate shift [48] (i.e., the alteration of output distribution caused by activation functions in the training process). In the proposed network, the batch normalization layer is deployed between the convolutional layer and the activation layer. Addition layer adds multiple inputs using an element-wise addition operation. In [31], the addition layer is exploited in skip connections (so-called residual
connections) to overcome the vanishing gradient problem, which in turn improves the learning efficiency. Concatenation layer involves multiple inputs and then concatenates them along a specified dimension. This layer is commonly utilized to collect more diversified features from multiple preceding layers for enriching pattern recognition [49]. The principle operations of the element-wise addition and depth-wise concatenation layers are described in Fig. 8.

III. DEEP LEARNING FOR AMC

In the past decades, DL with FFNN, RNN/LSTM, and CNN architectures has achieved remarkable success in wide domains, from computer vision [50]–[52] to bioinformatics [53]–[58]. Based on the superiority of learning high-level features directly from raw input data and dealing with big data effectively, DL plays as the core technology for many automatic pattern analysis based applications. Besides, DL is currently being adopted to solve many challenging problems in communications [4], [59]–[63]. For modulation classification, numerous methods have been introduced by exploiting different DL architectures, such as FFNN, RNN/LSTM, and CNN, to improve classification performance in terms of accuracy and processing speed. The overall framework of DL-based AMC methods is presented in Fig. 9, where the pre-processing step with data transformation and re-construction is optional [24]. The input of FFNNs is usually a vector of handcrafted features, while RNNs/LSTMs and CNNs can process the raw signal data directly. Moreover, with the good ability of extracting the representational features from high-dimensional unstructured data, the input of CNNs for modulation classification applications can be images, such as constellation diagram and spectrogram images. The performance of modulation classification is usually measured by the accuracy metric, which can be generally calculated as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN},
\]

where TP, TN, FP, and FN are the true positive, true negative, false positive, and false negative, respectively. Besides, the detailed results can be reported via confusion matrix, which is commonly used in the machine learning domain. This section overviews state-of-the-art DL-based AMC approaches, wherein several deep architectures are exploited for performance improvement.

A. FFNN-BASED AMC

Several AMC methods have been proposed by leveraging FFNN architectures to learn the handcrafted features. In [64], Kim et al. calculated numerous advanced features in the time and frequency domains to train a FFNN-based classifier, where the architecture of FFNN was mainly specified by three hidden layers having the numbers of neurons [500, 200, 40] and associating with the ReLU activation layers alternately. The designed FFNN was trained by using the stochastic gradient descent with momentum (SGDM) optimizer and the negative-log-likelihood to calculate loss function. Based on simulation results under the Rician channel with AWGN and high Doppler, the method outperformed the conventional approach using ANN with shallow architecture [65]. This FFNN-based approach was improved in [66] with feature selection that evaluates mutual information to select the best subset of features to reduce the complexity of model learning. In [67], an effective feature selection algorithm based on mutual information and correlation coefficient was proposed to reduce the number of features before passing to a FFNN having three hidden layers with the numbers of neurons [40, 20, 10]. The above-mentioned methods were evaluated for five modulation formats: BPSK, QPSK, 8-PSK, 16QAM, and 64QAM.

Some FFNNs have been proposed with specific layer configurations to deal with challenging issues or to improve classification accuracy. To flexibly identify data type, modulation class, and modulation order of the received signals, Karra et al. [68] deployed a hierarchical classifier that includes multiple FFNNs to deal with different formats of digital and analog modulations, such as single-sideband (SSB), double-sideband (DSB), and frequency modulation (FM). In [69], Xie et al. developed a FFNN-based digital modulation classifier with the input of high-order statistics. The FFNN was configured with four hidden layers associated with ReLU activation function for nonlinearity and cross-entropy function for weight updating strategy. With the number of neurons [5, 13, 6, 6] defined in four hidden layers, the network achieved high accuracy with six effortless low-order digital modulations, including ASK, FSK, and PSK. Similarly, Shi et al. [70] exploited a two-hidden-layer FFNN to learn higher-order cumulant features. To deal with more challenging high-order digital modulations, particle swarm optimization (PSO) algorithm was leveraged in hidden layers to optimally determine the number of nodes and further overcome the local minimum trap during the process of network training. In simulations, the numerical results demonstrated that PSO is more superior compared to the genetic algorithm (GA) in optimizing FFNN architecture.

Some advanced structures and sophisticated network designs have been recommended to adapt with various channel and noise conditions. An unsupervised modulation classification method [71] was introduced with an autoencoders based FFNN having two hidden layers. The synthetic data (including the real and imaginary samples of the received complex signal) was re-constructed into a high-dimensional array before feeding to the network for recognizing modulation patterns. Despite being more superior than many likelihood-based and feature-based approaches, this method was vulnerable with higher-order digital modulations under
TABLE 1. Summary of FFNN-based AMC methods.

| Year | Ref. | System | Modulation and Channel | Description/Highlight | HighAcc (SNR) |
|------|------|--------|------------------------|-----------------------|--------------|
| 2016 | [64] | SISO-SC | BPSK, QPSK, 8-PSK, and 16/64QAM | Statistical features in time-frequency domains | 100% (10 dB) |
|      |      |        | Rician channel, AWGN, and Doppler | A network with three hidden layers | |
| 2017 | [68] | SISO-SC | ASK, FSK, BPSK, QPSK, SSB, DSB, and FM | Raw signal input with IQ samples | 92.0% (18 dB) |
|      |      |        | AWGN, phase offset, and frequency offset | Hierarchical classifier with multiple FFNNs | |
| 2017 | [71] | SISO-SC | 2PAM, 4PAM, 8PAM, BPSK, QPSK, 8-PSK, 16/64/256QAM | High-dimensional input data array of real and imaginary | 100% (15 dB) |
|      |      |        | Flat fading, AWGN, and frequency offset | Autoencoder-based network with two hidden layers | |
| 2017 | [72] | MIMO-SC | BPSK, QPSK, 8-PSK, and 16QAM | Higher-order cumulants | 99.9% (15 dB) |
|      |      |        | Quasi-static Rayleigh fading and AWGN | Combination of FFNN and RBFN | |
| 2017 | [66] | SISO-SC | BPSK, QPSK, 8-PSK, and 16/64QAM | Statistical features and higher-order cumulants | 85.0% (5 dB) |
|      |      |        | Rayleigh fading, AWGN, and Doppler | Feature selection based on mutual information | |
| 2019 | [69] | SISO-SC | 2FSK, 4FSK, BPSK, QPSK, 2ASK, and 4ASK | High-order statistics | 100% (−2 dB) |
|      |      |        | AWGN | A network with two hidden layers | |
| 2019 | [70] | SISO-SC | BPSK, QPSK, 8-PSK, 16/64/256QAM | Statistical features and higher-order cumulants | 100% (12 dB) |
|      |      |        | AWGN | FFNN with PSO for optimizing weight updating | |
| 2020 | [73] | SISO-SC | BPSK, QPSK, 8-PSK, and 16QAM | Re-ordering signal samples | 98.0% (20 dB) |
|      |      |        | AWGN and frequency offset | FFNN with the MAP function | |
| 2020 | [67] | SISO-SC | BPSK, QPSK, 8-PSK, and 16/64QAM | Learning higher-order cumulants | 99.9% (10 dB) |
|      |      |        | AWGN | Selected features of mutual information and correlation | |

HighAcc: Highest accuracy, SISO-SC: single-input single-output single-carrier, MIMO-SC: multiple-input multiple-output single-carrier.

B. RNN/LSTM-BASED AMC

Rajendran et al. [76] proposed a data-driven model for AMC by exploiting a LSTM network to process the time domain a frequency-selective multipath Rayleigh fading channel. In [72], Shah et al. designed two relatively simple neural network architectures for a cost-efficient AMC method in MIMO systems. In particular, a sparse autoencoders based FFNN and a radial basis function network (RBFN) were deployed to process the instantaneous and higher-order cumulants extracted in the frequency domain of the received signals, in which the limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm and the least square mechanism were used to optimize the weight updating process. Although two networks were designed to be flexible with different MIMO configurations, these were quite sensitive to noise uncertainty. To deal with uncertain noise conditions, a blind modulation classification method [73] was introduced in single-input single-output (SISO) systems by exploiting a FFNN-based classifier incorporated with the maximum a posteriori probability (MAP) function. When the channel model becomes more complex and the number of modulation formats increases, re-ordering signal samples along the phase component in a pre-processing step can accelerate the training speed of FFNN. Based on performance evaluation, this method revealed two practical issues: the feasibility of FFNN under channel impairments and the robustness of classification model with higher-order formats. In a nutshell, we summarize existing FFNN-based AMC methods in Table 1, wherein the information of target modulations, channel conditions, and principal techniques of methods are given in brief.

TABLE 2. RadioML2016.10A dataset [74].

| Number of modulations | 11 |
|-----------------------|----|
| Modulation formats    | BPSK, QPSK, 8-PSK, 16QAM, 64QAM, 4PAM, GFSK, CPFSK, WBFM, AM-SSB, AM-DSB |
| Signal length         | 128 complex samples |
| SNR range             | −20 : 2 : 20 dB |
| Total number of signals | 165,000 |
| Channel specification | Selective multipath Rician fading |
|                       | Sample rate offset |
|                       | Center frequency offset |
|                       | AWGN |
Some advanced RNN architectures have been introduced by combining with FFNNs and incorporating with other processing units to improve learning efficiency. In [88], a novel hierarchical RNN architecture was designed with a grouped auxiliary memory module to overcome the vanishing gradient problem and also capture long-term dependencies effectively. This module can effectively maintain the information from previous time steps to drive a new state at each hidden layer, thus revealing more temporal correlations to improve correct identification rate. Furthermore, this memory mechanism can facilitate different unit types, including naive RNN, LSTM, and GRU. In [87], Ghasemzadeh et al. introduced a novel AMC method with a two-fold contribution: (i) the higher-order cumulants and polar coordinates of IQ samples were extracted from underwater acoustic signals, improve recognition capability, and reduce computational complexity. To handle variable length signals, a masking mechanism in the pre-processing stage added padding zeros automatically to the input. This hybrid network was trained with the Adam optimizer and the MSE-based cross entropy function. The BiGRU-based network can achieve cost efficiency and prevent the overfitting problem more effectively than several regular RNN and LSTM networks.

Some RNN-based methods have adopted GRU units instead of LSTM units in hidden layers to achieve cost efficiency. Hong et al. [81] designed a RNN architecture having two GRU layers to seize the temporal sequence characteristics of received signals. The outcome of the second GRU layer were fed into a fully connected layer with the softmax function for modulation classification. In the training stage, the network was trained with a root mean square propagation (RMSProp) optimizer with categorical cross entropy function. To evaluate the parameter sensitivity, the network was tested with different numbers of hidden layers and different numbers of GRU units. In [82], a cost-efficient RNN was developed with one GRU layer followed by a fully connected layer to identify modulations in resource-constrained IoT devices. Sun and Wang [83] proposed a cooperative and time-varying bidirectional RNN for wireless signal detection, wherein hidden layers can be configured with naive RNN, LSTM, and GRU units in an one alternate manner. The meaningful features of different hidden layers were associated at multiple time steps over an intermediate time-varying merge layer, and the features of the last layer in forward and backward paths were gathered at a fully connected layer subsequently.

High-level versions of LSTM and GRU layers, including bidirectional LSTM (BiLSTM) and bidirectional GRU (BiGRU), have been studied to enhance learning efficiency. In [84], Daldal et al. designed a lightweight LSTM network with one BiLSTM layer [85] and one fully connected layer in the architecture to identify modulations under noise condition. As shown in Fig. 10, BiLSTM consists of two LSTMs stacked along the forward and backward paths, which can induce the output as the combination of the two LSTMs. BiLSTM can train and reverse the time series data to fully reveal the information correlations of the front and back signal sequences. In the simulations with three modulation classes (e.g., ASK, PSK, and FSK), the BiLSTM-based classifier performed modulation identification more precisely than SVM classifier at different SNR levels. Wang et al. [86] designed a hybrid time series network with the architecture involving multiple GRU and BiGRU layers (see Fig. 10b) to obtain three-fold objective: enrich underlying features extracted from underwater acoustic signals, improve recognition capability, and reduce computational complexity. To handle variable length signals, a masking mechanism in the pre-processing stage added padding zeros automatically to the input. This hybrid network was trained with the Adam optimizer and the MSE-based cross entropy function. The BiGRU-based network can achieve cost efficiency and prevent the overfitting problem more effectively than several regular RNN and LSTM networks.

### TABLE 3. RadioML2018.01A dataset [75].

| Number of modulations | Analog Modulations | Digital Modulations |
|-----------------------|--------------------|---------------------|
|                       | FM, AM-SSB-SC, AM-SSB-WC, AM-DSB-SC, AM-DSB-WC | 8-PSK, QPSK, 16-PSK, 32-PSK, 64-PSK, 128-PSK, 256-PSK, 1QAM, 2QAM, 4QAM, 8QAM |
| Signal length         | 1024 complex samples |
| SNR range             | $-20 : 2 : 30$ dB |
| Total number of signals | 2,555,904 |
| Channel specification | Selective multipath Rician fading, Carrier frequency offset, Symbol rate offset, Non-impulsive delay spread, Doppler shift, AWGN |

radio peripheral (USRP) testbed scenarios. An accurate LSTM-based method was introduced in [78] to effectively deal with various channel impairments (Rayleigh fading and mixture AWGN) by additionally exploiting a temporal attention layer to enrich features. Based on the simulation results of four digital formats (e.g., BPSK, QPSK, 8-PSK, and 16QAM), the network proved the superiority over traditional classifiers and other regular LSTM networks.

Chen et al. [79] proposed a single-layer LSTM network with a random erasing-based test time augmentation mechanism (RE-TTA) to achieve cost efficiency without sacrificing classification accuracy. By associating an attention module between the LSTM layer and the fully connected layer, the spatial correlations between different AP samples in an input signal were extracted to improve classification accuracy. Based on the numerical results obtained by evaluating on the RadioML2016.10A dataset, the proposed method proved the effectiveness of RE-TTA and outperformed some basic RNN and LSTM based approaches in terms of classification accuracy. The work [80] adopted different data augmentation methods to overcome the overfitting problem caused by insufficient training data. Concretely, by analyzing the radio statistics of received signals, three augmentation methods were utilized for both the IQ and AP samples in the training and prediction stages of a LSTM-based classifier.

Based on the numerical results obtained by evaluating on the RadioML2016.10A dataset, the proposed method proved the effectiveness of RE-TTA and outperformed some basic RNN and LSTM based approaches in terms of classification accuracy. The work [80] adopted different data augmentation methods to overcome the overfitting problem caused by insufficient training data. Concretely, by analyzing the radio statistics of received signals, three augmentation methods were utilized for both the IQ and AP samples in the training and prediction stages of a LSTM-based classifier.
and low execution latency, respectively. Interestingly, both the DBN and SNN were leveraged as the fully-connected network (FCN) in the platform, which played the role of classification with different learning strategies. Based on the numerical results obtained on the RadioML2018.01A dataset [75] (where the detailed description is provided in Table 3), DBN improved the average accuracy by around 16% and SNN reduced the execution latency by over 34% comparing with the standard RNN.

RNN architectures were exploited for modulation classification of radar pulse repetition interval (PRI) signals [89] and acoustic signals [90]. Although traditional PRI modulation classification approaches can deal with simple PRI modulations, their performance is dramatically decreased under channel deterioration. Li et al. [89] proposed a high-performance method to recognize several complex PRI modulations of radar signals by exploiting RNN architecture, wherein GRU units were incorporated with an attention unit to solve the problem of high ratios of lost and corrupted pulses. Yu et al. [90] adopted a basic LSTM network having one hidden layer to recognize modulation patterns of underwater acoustic non-cooperative communication signals. By mining the instantaneous spectrum characteristics of acoustic signals, this network was able to deal with some specific modulation techniques, such as OFDM and direct-sequence spread spectrum.

Fig. 11 presents the overall architectures of some state-of-the-art deep RNNs, including Rajendran et al. [76], Wang et al. [86], Hu et al. [78], and Ghasemzadeh et al. [87]. Such architectures were proposed to increase accuracy and reduce computational cost besides handling some challenging issues, such as unfixed-length signal and vanishing gradient. We summarize state-of-the-art RNN-based methods in Table 4. In general, RNN and LSTM are more superior than FFNN thanks to the good ability of LSTM and GRU units to extract temporal dependencies of signal. Besides, the learning
TABLE 4. Summary of RNN/LSTM-based AMC methods.

| Year | Ref. | System | Modulation and Channel | Description/Highlight | HighAcc (SNR) |
|------|------|--------|------------------------|-----------------------|--------------|
| 2017 | [81] | SISO-SC | RadioML2016.10A dataset | RNN with two GRU and one fully connected layers Training with the RMSProp optimizer | 91.0% (10 dB) |
| 2018 | [76] | SISO-SC | RadioML2016.10A dataset | Two LSTM layers and one fully connected layer Processing the averaged magnitude FFT data | 93.0% (20 dB) |
| 2018 | [78] | SISO-SC | BPSK, QPSK, 8-PSK, and 16QAM Rayleigh fading and mixture Gaussian noise | Three LSTM layer and four fully connected layers Feature enrichment with a temporal attention layer | 98.0% (20 dB) |
| 2018 | [87] | SISO-SC | BPSK, QPSK, 128QAM, and 256QAM AWGN, multipath fading, and frequency offset | Higher-order cumulants and polar coordinates Two LSTM layers associated with DBN and SNN High accuracy and low latency | 100% (30 dB) |
| 2019 | [77] | SISO-SC | BPSK, QPSK, 8PAM, 16/64/128/256QAM Multiple TDL channel models with AWGN | LSTM adapts the sub-Nyquist sampling rates Performance validation on USRP testbed Being robust under various channel impairments | 98.0% (25 dB) |
| 2020 | [79] | SISO-SC | RadioML2016.10A dataset | Spatial attention module Random erasing-based augmentation scheme | 94.1% (18 dB) |
| 2020 | [80] | SISO-SC | RadioML2016.10A dataset | Two LSTM and two fully connected layers Three data augmentation methods based on performing rotation, flip, and noise addition | 92.0% (16 dB) |
| 2020 | [82] | SISO-SC | Practical extension of RadioML2016.10A Multipath fading, AWGN, and frequency offset | Specific designed for resource-constrained devices Lightweight RNN with single GRU-based hidden layer | 93.1% (22 dB) |
| 2020 | [86] | SISO-SC | BPSK, QPSK, 8-PSK, PAM, 16/32QAM, CPFSK, GFSK, FM, and SSB Shallow sea model with AWGN | Three-fold objective: feature enrichment, improve recognition capacity, and reduce complexity Hybrid time series network with BiGRU layers and zero masking | 94.0% (20 dB) |
| 2020 | [88] | SISO-SC | RadioML2016.10A dataset | Grouped auxiliary memory module Reveal long-term dependencies of unixed-length signals | 92.7% (20 dB) |

efficiency of RNN and LSTM can be improved by some supplementary components, such as spatial attention layer and grouped auxiliary memory module.

C. AMC WITH CNNS
1) SEQUENTIAL signal-based approaches
Besides introducing the RadioML2018.01A dataset with 24 formats (including analog and digital modulations), O’Shea et al. [75] deployed two CNNs by adapting the architectures of VGG and ResNet with one-dimensional asymmetric convolution filters in layers. Based on the simulation results obtained on the dataset, two CNNs significantly outperformed the FB method thanks to the deep features learned by convolutional layers, and ResNet performed classification more accurately compared to VGG because of residual blocks shown in Fig. 12a to prevent networks from the vanishing gradient problem. Furthermore, two CNNs were investigated with different signal lengths and different numbers of training samples under a synthetic channel impairment, involving multipath fading, carrier frequency offset, symbol rate offset, and AWGN. To facilitate the varying input dimensions, a novel CNN with two training stages [91] was deployed for classifying the modulations of long symbol-rate observation sequences. The proposed CNN, namely CNN-AMC, incorporated the raw signal data with the estimation of symbol SNR as the supplemental information to enhance pattern learning. The first stage (pre-training) trained CNN-AMC on a basic dataset in the presence of AWGN and the second stage (fine-tuning) tuned the trained model with a new dataset by replacing some top layers with random parameter initialization. As a result, this training strategy can deal with different channel conditions and adapt various modulation patterns (i.e., being easy to update model with a new dataset). For performance evaluation, several simulation results were provided to show the effectiveness of two-training strategy and also the robustness under channel deteriorations. Besides, the proposed CNN-AMC revealed the preeminence in terms of accuracy and complexity over some existing FB approaches. Numerous CNNs have specified sophisticated convolutional blocks and advanced processing modules by cleverly incorporating multiple convolutional layers with others operation layers to improve diversified features. In [92], a cost-efficient and well-performed CNN, namely MCNet, was proposed for robust modulation identification under various channel impairments. This network was designed with different 1D asymmetric filters in a convolutional block, as shown in Fig. 12b, to significantly reduce the network size (i.e., the total number of trainable parameters) without sacrificing accuracy. Additionally, the concatenation layers were adopted to collect more diversified features of multiple convolutional layers and the element-wise addition layer was deployed in skip connections to prevent MCNet from vanishing gradient, thus improving the learning efficiency. Based on performance evaluation on the RadioML2018.01A...
dataset, MCNet outperformed the adapted VGG and ResNet in [75] while keeping a lower network complexity. The strategy of using multiple convolutional layers with different symmetric filter sizes was also adopted in [95]. Huynh-The et al. [96] proposed a novel reusable-feature CNN as a high-accuracy modulation classifier for cognitive radio in wireless communications. In additions to the skip connections to prevent vanishing gradient, a reusable-feature module with depth-wise concatenation was developed to optimize the feature utilization by gathering deep features at multi-scale representations. The max pooling layers with different pool sizes and strides were arranged to align the different spatial sizes of module’s outcomes before feature aggregation.

Optimally learning the underlying features of modulated signals at multi-scale resolutions was done by several processing blocks having the connection structure inspired from the shape of catenary. These blocks were organized by following a cascade connection in a novel CNN, namely Chain-Net [93]. Each block involved two parallel convolutional flows, as shown Fig. 12c, was associated via a depth-wise concatenation layer, where each flow was specified by 1D asymmetric convolution filters to compute the locally temporal relation between IQ samples and the cross-component correlation of each sample. In [94], Tunze et al. combined different types of convolutional layers (e.g., regular, grouped, and depth-wise) in a sparsely connected CNN (SCGNet) to improve accuracy while keeping a low complexity. SCGNet includes a generic feature extraction module to calculate coarse features, a speed-accuracy trade-off module to enrich the relevant features of preceding layers, and a deep feature extraction and processing module to reveal more discriminative features. Fig. 12d illustrated the structure of deep feature extraction and processing module with several cascaded multi-dimensional grouped convolutional units. Zhang et al. [97] introduced a learning modulation filter network (LMFN) with a two-stage optimization algorithm. The modulation filters in LMFN which were designated to enhance the learning capacity of filters associated with attention modules to selectively collect the meaningful properties of input signals. The work [98] proposed modulated autocorrelation convolution networks (MACNs), wherein the modulation filters performed a novel autocorrelation convolution to capture periodic characteristics of received signals. MACNs achieved acceptable accuracy with a small model size for being suitable to storage constrained devices. To deal with crashed signals due to additive noise, a novel CNN-based framework [99] was introduced with three modules: SNR prediction, classification, and signal processing, in which the input signals estimated with low SNR were first re-constructed by U-Net [100] for signal re-construction and enhancement and then provided to a CNN for classification. To handle the unknown intermediate frequency and non-cooperative modulation, Liu et al. [101] introduced a novel frequency selection layer to first detect the frequency band of signals and then filter out the out-of-band noise. In [102], the problem of unfixed-length signals feeding to the fixed-size input layer of CNNs was solved by three fusion mechanisms, including feature-based fusion, confidence-based fusion, and voting-based fusion, which were integrated into the adapted multi-stream architectures of VGG and ResNet [75]. Gu et al. [103] designed a plain-architecture network, namely blind channel identification aided generalized automatic modulation recognition (BCI-AMR), with multiple dropout layers [104] to effectively prevent the network from overfitting.

Several CNNs have utilized various techniques, such as fine-tuning and pruning [105], to reduce the computational complexity of deep networks to accelerate processing speed. Some AMC methods [106]–[108] shared a same simple architecture to recognize few modulation formats, including BPSK, QPSK, and 8-PSK. The pruning and fine-tuning techniques were cooperatively applied in [108], [109] to reduce computational complexity. In the simulations, although this work significantly reduced the network size and computation time, the classification results on few low-order digital modulation formats was unremarkable. Wang et al. [110] extended the work [108] for classifying modulations in MIMO systems. To deal with the unknown SCI scenario, a channel estimation algorithm was incorporated with a zero-forcing equalization algorithm to increase SNR of the received signals, thus improving the classification accuracy in CNN. Compared with a conventional method adopting high-order cumulants and ANN, the proposed method was more superior for different numbers of transmitter and receiver antennas under both known and unknown CSI conditions. Besides the deployment of CNNs for learning intrinsic features of MIMO signals, a decision maker [111] with various cooperative

![FIGURE 12. Primary components: (a) residual block in an adapted ResNet [75], (b) m-block in MCNet [92], (c) chain-block in Chain-Net [93], and (d) deep feature extraction and processing module in SCGNet [94].](image-url)
rules, such as direct voting, weighty voting, direct averaging, and weighty averaging, was developed to improve accuracy. Among four fusion rules, the weighted average ones exhibit the highest performance by summarizing the underlying correlations between signals of different antennas effectively. Hong et al. [112] upgraded the work [103] for OFDM systems. Although the proposed method performed modulation identification more accurately than conventional FB approaches with higher-order cumulants and different classification algorithms, the advantage of CNNs for OFDM signals was not clarified convincingly comparing with regular single-carrier signals. It was observed that the architecture of CNNs in above-mentioned AMC approaches [103], [106]–[108], [110], [112] were not optimized to effectively learn intrinsic features in multi-scale representations, that means, the shallow features of two convolutional layers are vulnerable under channel impairments to inappropriately classify higher-order digital modulations. Besides, another limitation was a huge number of trainable parameters, such as the number of parameters of BCI-AMR [103] dramatically increases in fully connected layers. This drawback can be overcome by arranging the max pooling layers and the global average pooling layers to reduce the number of parameters in fully connected layers and also the computational cost in convolutional layers.

In some practical scenarios, data distribution is nonuniform due to different sampling rates, and therefore, a classification model trained on a source dataset may not be capable of representing a target dataset. To overcome this challenge, Bu et al. [113] introduced an adversarial transfer learning architecture (ATLA), wherein two stages of adversarial training and knowledge transferring were performed by two CNNs and incorporated in an unified learning framework to improve the accuracy of modulation classification. Based on the simulation results obtained on RadioML2016.01A, the ATLA-based learning framework demonstrated its superiority over existing transfer learning algorithms in terms of transfer capability and the prevention of dataset bias. In [114], Zhou et al. proposed to use capsule networks [115] for blind modulation classification to address the problem of overlapped co-channel signals in a system with multiple transmitters. By specifying multiple capsule layers, the representational features of regular convolutional layers were enriched to distinguish modulation patterns accurately. Moreover, Li et al. [116] built a novel network architecture with primaryCaps and digitCaps derived from capsule layers, namely AMR-CapsNet, to solve the problem of limited training data in practice. In [117], a waveform-spectrum multimodal fusion approach classified modulations by merging the deep features which are extracted from multiple deep residual networks with different data modalities, such as IQ, AP, and spectrum. In Table 5, we summarize state-of-the-art CNN-based methods which enable the raw sequential signals as the input of networks. Some advanced layer structures with skip connection and feature concatenation are recommended to prevent vanishing gradient and enrich diversified features.

2) IMAGE-BASED APPROACHES

Apart from raw signal data with IQ and AP samples, the constellation image and spectrum image of digital modulations have been utilized to automatically identify modulations by CNNs [120], where the modulation classification is regarded as an image classification task (see a general workflow in Fig. 13). In [121], Peng et al. evaluated the classification accuracy of AlexNet [122] and GoogleNet [123] on different datasets of gray-scale and color constellation images. Moreover, the performance of CNNs was investigated with various image resolutions and network configurations. Although this approach performs modulation classification more accurately than the FB methods using the cumulants and SVM classifier, it is more complex and requires a longer processing time. To reduce the network complexity, Huang et al. [124] proposed a lightweight CNN with few convolutional layers to capture the representational features from regular constellation images and contrast enhanced grid constellation images. Moreover, this network exploited intra-class compactness and inter-class separability using a compressive loss constraint to improve the accuracy of higher-order digital modulation. Wang et al. proposed a hierarchical CNN-based modulation classifier [125], in which one CNN was designed to classify low-order digital modulation formats using IQ samples and another one was specified to discriminate high-order digital modulation formats by learning visual features from constellation images. In the above-mentioned CNNs, the activations resulted by the last convolutional layer were flattened to directly connect with neurons in the first fully connected layer, which dramatically increases the number of trainable parameters. This poor design strategy can be addressed by arranging a global average pooling layer before the first fully connected layer to reduce the number of parameters and also prevent overfitting.

Some deep CNNs have been designed with various advanced layer-wise structures to optimize visual feature learning efficiency. The work [126] proposed a cross-residual CNN, namely CRNet, for constellation image-based modulation classification, which mainly consisted of two parallel convolutional flows associated via a novel cross-residual mechanism to enrich the deep visual features. To effectively handle the problem that multiple scattered points are quantized to a pixel coordinate without the notice of power and position of symbols, Doan et al. [127] performed a bivariate histogram equalization and an exponential decay mechanism to calculate the pixel values in a constellation image. Huang et al. [128] proposed an efficient CNN to learn a contrastive loss function to train CNNs for a grid constellation.
TABLE 5. Summary of CNN-based AMC methods with sequential signals.

| Year | Ref. | System | Modulation and Channel | Description/Highlight | HighAcc (SNR) |
|------|------|--------|------------------------|------------------------|---------------|
| 2018 | [75] | SISO-SC | RadioML2018.01A dataset | Introduction of a new dataset | 95.0% (20 dB) |
| 2018 | [91] | SISO-SC | PSK, QPSK, 8-PSK, and 16/32/64QAM Flat Rayleigh fading channel with AWGN | CNN with two training stages with known SNR | 100% (6 dB) |
| 2019 | [102] | SISO-SC | BPSK, QPSK, 8-PSK, OQPSK, 2-PSK, 4-PSK, 8-PSK, 16/32/64QAM, 4PAM, and 8PAM AWGN, carrier phase and frequency offset | Deployment of three fusion techniques: feature-based, confidence-based, and voting-based fusion | 99.9% (30 dB) |
| 2019 | [114] | SISO-SC | 2-ASK, 2/4-PSK, BPSK, QPSK, 16/64QAM, and MSK AWGN and frequency offset | Exploiting capsule networks to address overlapped co-channel signals | 95.0% (9 dB) |
| 2020 | [92] | SISO-SC | RadioML2018.01A dataset | MCNet with various 1D asymmetric filters | 93.6% (20 dB) |
| 2020 | [109] | SISO-SC | RadioML2016.10A dataset | Network pruning with absolute sum and APoZ pruning | 81.4% (4 dB) |
| 2020 | [106] | SISO-MC | BPSK, QPSK, and 8-PSK Rayleigh channel, AWGN, and carrier phase offset | Two convolutional and three fully connected layers | 100% (10 dB) |
| 2020 | [108] | SISO-SC | BPSK, QPSK, 8-PSK, and 16QAM Time-invariant channels, flat fading, and AWGN | Channel estimation with zero-forcing equalization | 100% (10 dB) |
| 2020 | [110] | MIMO-SC | RadioML2018.01A dataset | Grouped and separable convolutional layers | 94.4% (20 dB) |
| 2020 | [93] | SISO-SC | FM, 16/32/64/128QAM, 16/32/64/128APS, 16PAM, AM-DSB-WC, AM-DSB-SC, AM-SSB-WC, and AM-SSB-SC Multipath fading, AWGN, and Doppler | Multiple chain blocks connected in a cascade | 85.3% (20 dB) |
| 2020 | [113] | SISO-SC | RadioML2016.10A dataset | Adversarial transfer learning architecture | 81.7% (18 dB) |
| 2020 | [97] | SISO-SC | 3/4/8-ASK, 2/4/8-PSK, BPSK, QPSK, 8-PSK, and 64/256QAM Multipath Rayleigh fading channel with AWGN | Two-stage optimization algorithm | 88.8% (20 dB) |
| 2021 | [117] | SISO-SC | RadioML2016.01A dataset | Waveform-spectrum multimodal fusion mechanism | 99.0% (30 dB) |
| 2021 | [116] | SISO-SC | RadioML2016.01A dataset | Exploiting capsule networks to overcome data scarcity | 90.0% (12 dB) |
| 2021 | [118] | MIMO-SC | BPSK, QPSK, 8-PSK, and 16QAM Rayleigh channel, AWGN and carrier phase offset | Federated and decentralized learning with CNNs | 98.0% (10 dB) |
| 2021 | [119] | SISO-SC | BPSK, QPSK, 8-PSK, and 16QAM Rayleigh channel, AWGN, and carrier phase offset | Dealing with class imbalance and noise varying | 100% (20 dB) |

matrix-based method, which allows the deep network to increase the discrimination among different modulation patterns to further boost classification accuracy. In [129], a cost-efficient learning mechanism for AlexNet was recommended to classify modulations from constellation diagrams, which involved three processing stages: training from scratch, average percentage of zeros (APoZ) pruning [130], and fine-tuning. The proposed approach was experimented on the NVIDIA Jetson TX2 module (an AI computing device) for feasibility measurement besides simulations on an artificial dataset. Huang et al. [131] proposed an AMC method to identify physical-layer attacks for IoT security. This method was leveraged by a multi-module fusion CNN to process constellation diagrams generated by a pixel-coloring...
constellation projection algorithm. In [132], the cyclic spectrum (CS) image and constellation diagram (CD) image of a modulated signal were learned separately by two convolutional streams, wherein the high-level features resulted by two streams were fused to enrich feature diversity. This approach outperformed [76] on the RadioML2016.10A dataset while presenting a cheaper computational cost. Zeng et al. [133] connected four convolutional layers with three max pooling layers alternately in cascade to process the spectrogram images obtained by the short-time discrete Fourier transform. As the pre-processing step, low-pass filters were applied to spectrogram images for noise removal before feeding to the network. Table 6 summarizes state-of-the-art CNN-based methods which are specially designed to process the constellation images and diagrams of modulation signals.

D. HYBRID DL MODELS

Some works have taken advantages of RNN (i.e., capable of mining long-term correlations of signals with adaptive length) and CNN (i.e., a comprehensive learning capability of representational features in multi-scale resolutions) to improve the overall modulation classification of systems. For example, Huang et al. [135] proposed a novel gated recurrent residual network (GrrNet) to identify modulations precisely. GrrNet with the overall architecture given in Fig. 14a primarily consisted of three modules: ResNet-based feature extraction, feature fusion, and GRU-based classification. The locally intrinsic features, extracted by convolutional layers at residual blocks, were flatten into vectors which were than fed into GRU modules to calculate the temporal correlations between locally neighboring time steps and also from all the preceding steps. This processing strategy was flexible to different input signal lengths beside classification performance improvement. Based on the simulation results of five modulation formats {BPSK, QPSK, 8-PSK, 16QAM, 64QAM}, GrrNet obtained the average accuracy of approximately 99% at 5 dB SNR and outperformed ResNet [75] and LSTM [76] significantly at multiple SNR levels. The same network design strategy was presented in HybridNet [136] with three convolutional blocks followed by BiGRU blocks and fully connected layers as shown in Fig. 14b. Besides the residual connection, each block was specified by a squeeze and excitation (SE) block to maximize the intra-class discrimination. Notably, the authors adopted two connection branches, where each branch consists of one BiGRU and two fully connected layers, to reveal the temporal dependencies explicitly and enrich the feature diversity. In the performance evaluation on a dataset with eight digital modulations and two analog modulations, the proposed network obtained the average accuracy of over 93% at SNR $\geq 0$ dB (when training with 50% dataset size) and also demonstrated the effectiveness of BiGRU in terms of generalizing deep models using a small number of training samples beside the advantage of SE blocks in increasing the accuracy of higher-order modulations.

To simultaneously acquire the deep features from the IQ and AP samples of received signals, Zhang et al. [137] deployed a deep network with dual-stream architecture, namely CNN-LSTM, where each stream consisted of three convolutional units (groups of convolutional layers and ReLU layers) followed by two LSTM layers as presented in Fig. 14c. Two LSTM layers in each stream were capable of mining the long-term dependencies as temporal correlations based on computing the outcomes at the input gate and forget gate. By interacting the outputs of two streams in pairs via a multiplication operation, the diversity of features was increased significantly. Based on the simulation results on the RadioML2016.10A dataset, the proposed CNN-LSTM proved the effectiveness of incorporating two processing
TABLE 6. Summary of CNN-based AMC methods with images.

| Year | Ref. | System                  | Modulation and Channel                                      | Description/Highlight                                                                 | HighAcc (SNR)   |
|------|------|-------------------------|------------------------------------------------------------|----------------------------------------------------------------------------------------|-----------------|
| 2019 | [121] | SISO-SC                 | PSK, QPSK, 8-PASK, OQPSK, 4-ASK, and 16/32/64QAM AWGN       | Converting to constellation images Image classification with AlexNet and GoogleNet       | 100% (10 dB)    |
| 2019 | [124] | SISO-SC                 | PSK, QPSK, 8-PASK, and 16/64QAM Multipath Rayleigh fading, AWGN, and frequency offsets | CNN with a compressive loss constraint Constellation images and contrast enhanced grid constellation images | 100% (15 dB)    |
| 2019 | [133] | SISO-SC                 | RadioML2016.10A dataset                                    | Transforming to spectrogram images Low-pass filter-based noise reduction                | 91.0% (10 dB)   |
| 2019 | [125] | SISO-SC                 | BPSK, QPSK, 8-PASK, GPSK, CPFSK, 4PAM, and 16/64QAM Multipath Rayleigh fading and AWGN | Two CNNs to process IQ samples and constellation images Highly computational cost       | 98.0% (18 dB)   |
| 2019 | [129] | SISO-SC                 | BPSK, QPSK, OQPSK, 8-PASK, 4-ASK, and 16/32/64QAM AWGN     | Network pruning and fine-tuning in constellation-input CNN Complexity reduction without sacrificing accuracy | 100% (6 dB)     |
| 2019 | [128] | SISO-SC                 | BPSK, QPSK, 8-PASK, and 16/64QAM AWGN                     | Transformation of grid constellation matrix Contrastive loss function to increase the pattern discrimination | 100% (15 dB)     |
| 2020 | [131] | SISO-SC                 | PSK, QPSK, 8-PASK, 4PAM, 16/64QAM Rayleigh channel, AWGN, and frequency offset | Multi-module fusion CNN Pixel-coloring constellation projection algorithm              | 100% (14 dB)    |
| 2020 | [126] | SISO-SC                 | QPSK, 8-PASK, 4/16PAM, 16/32/64APS, and 16/32/64QAM Multipath Rayleigh fading, AWGN, and Doppler effect | Bivariate histogram process for gray-scale constellation image generation Deployment of cross-residual mechanism in each processing module | 100% (20 dB)    |
| 2020 | [127] | SISO-SC                 | QPSK, 8-PASK, 4/16PAM, 16/64APS, and 16/32QAM Multipath Rayleigh fading and AWGN | Exponential decay mechanism to overcome the overlapping issue Deployment of Flow-in-Flow blocks | 100% (10 dB)    |
| 2021 | [120] | SISO-SC                 | BPSK, QPSK, 8-PASK, and 16/64QAM                          | Gray-scale constellation image generation CNNs with skip-connection                  | 100% (10 dB)    |
| 2021 | [134] | SISO-SC                 | RadioML2016.10A dataset                                    | Combining LSTMs for raw data and CNNs for constellation images Verification with different DL-based classifiers | 92% (18 dB)     |

Few recent methods have leveraged fusion mechanisms to combine the underlying radio features and visual representational features extracted from different network architectures. The work [140] introduced a convolutional and recurrent fusion network (CRFN) with four kinds of structure combination: (i) CNN series simple recurrent unit (SRU), wherein multiple convolutional layers were followed by a block of SRUs, (ii) SRU series CNN, wherein a block of SRUs was followed by a stack of convolutional layers, (iii) CNN parallel SRU, wherein two processing flows respectively specified by a convolutional block and a SRU block were gathered by a concatenation layer, and (iv) weighted average, wherein the prediction scores conducted by two flows were combined and averaged to make the final decision. To improve accuracy under imperfect channels, Zhang et al. [139] recommended a feature-based fusion framework (as shown in Fig. 15) to take the advantages of handcrafted features and deep features. The modulated signal was first transformed to the spectrum images using the time-frequency analysis (TFA) algorithm with Born-Jordan distribution (BJD) and smooth pseudo Wigner-Ville distribution (SPWVD) and then forwarded to the pre-trained ResNet to calculate the deep features. The handcrafted features from raw signals (including descriptive statistics and higher-order cumulants) were combined with the deep features of two spectrum images by a multi-modality feature-based fusion module before classifying with FFNN.

![FIGURE 15. Multi-modality feature-based fusion framework in [139] for modulation classification.](image-url)
Based on the performance evaluation with several low-order digital modulation formats, the proposed fusion framework reported the state-of-the-art accuracy while extremely increasing the complexity with two CNNs for visual feature extraction and one FFNN for classification beside two TFA algorithms for transforming spectrum image.

IV. RESEARCH CHALLENGES AND FUTURE DIRECTIONS

This section discusses the research challenges posed by DL-based AMC: signaling model, practicality, and performance balance. Accordingly, we deliver some promising future directions to develop a high-performance DL model for modulation classification in wireless communications.

A. SIGNALING MODEL

Several traditional and innovative AMC algorithms have been presented for single-carrier communication systems, whereas few methods have paid the attention on classifying modulations in multi-carrier systems, such as OFDM [141]–[143]. In future wireless communication systems, OFDM will be a widely used multi-carrier modulation technology, which has proved the improvement of spectrum utilization and the ability of dealing with multipath fading effectively. In [112], a DL-based method was introduced to identify the modulations in OFDM systems, however, without the enlightenment of the theoretical hypothesis, the motivation to deploy CNN for processing OFDM signals is questionable if compared with other CNN-based approaches for regular single-carrier signals. Recently, some AMC methods have exploited DL to identify modulations in MIMO systems [110], [111], but the superiority of deep networks to improve the overall accuracy was not discussed sufficiently. Another issue in dynamic spectrum access is burst signal (i.e., the received signal is featured by short duration with uncertain starting and ending time stamp) which poses some challenges of designing a deep model to deal with varying-length signals effectively. A universal AMC approach should be developed to deal with the diversified-format signals of different systems effectively (e.g., the inter-correlations between different OFDM symbols can be learned by deep models) and to adapt with varying-length signals. To this end, the recursive connection [144] is recommended to deploy in a hybrid RNN-CNN deep models.

In the era of 5G and IoT, the crowded spectrum environment of densely connected devices brings the multi-signal coexistence problem. In this harsh scenario, a target signal has to be first detected and identified accurately before recognizing its other characteristics, such as modulation type. Multi-task learning becomes a promising solution to recognize signals and classify modulation concurrently.

B. PRACTICALITY

Most of existing AMC methods have been evaluated on the simulation datasets which are generated by software. From the perspective of practicality, generating modulated signals to benchmark performance of deep models should take into account two primary concerns: modulation technique (including modulation types and the number of modulations) and channel condition. For examples, RadioML2018.01A, a currently available and widely used dataset of modulation classification, covers up to 24 modulations (including analog and digital techniques, and several challenging high-order formats), where the modulation signals are propagated in a multipath Rician fading channel with carrier frequency offset, symbol rate offset, delay spread, and AWGN to nearly obtain real-world phenomena in wireless communications. Numerous AMC methods have generated effortless datasets with few simple digital modulations. Consequently, the soundness and impact of research contribution based on performance evaluation in simulations can be unconvincing besides the unsubstantial reliability of classification models to implement in practical systems. For instance, a plain-architecture deep network [106] achieved high accuracy with three modulation candidates (BPSK, QPSK, and 8-PSK). In other works [108], [110], [111], the considered propagation channel is less challenging with flat fading and time-invariant instead of frequency-selective multipath fading and sampling time drift. Obviously, the modulation classification task becomes more effortlessly with a small number of given classes, where the inter-class and intra-class discrimination problems are not deliberated strictly. These approaches ignored higher-order digital modulation formats which are commonly used in high-speed communications.

The aforementioned issues encourage research communities to follow as much as possible the practical requirements regarding modulation format and propagation channel whenever an AMC algorithm is developed and evaluated to guarantee practicality. Additionally, most of the existing DL-based AMC approaches were studied for transmission scenarios with AWGN, while there are few traditional methods [145], [146] took into consideration in the $\alpha$—stable noise environment. It is worth noting that the noise in practical wireless systems usually presents impulse characteristics in the statistical theme, which can be modeled as an $\alpha$—stable distribution. Exploiting DL to accurately identify modulations under an $\alpha$—stable impulsive noise environment is promising and should be investigated in future.

C. PERFORMANCE BALANCE

Developing an effective AI-based approach to achieve a good performance in terms of accuracy and processing speed is always a complicated study, and becomes increasingly difficult with DL. Compared with traditional machine learning algorithms, DL consumes more computational resources in general. The implementation of deep model can be an awful experience with edge devices having limited computing capability, low memory and small storage space. In fact, designing a deep network architecture to applyably implement on the resource-constrained devices while attaining a comfortable performance requires solid knowledge in the field. For example, BCI-AMR was designed without the pooling layers in the architecture to reduce the dimensions of feature maps, that makes the following fully...
connected layers to load a huge number of parameters and a high computational cost. Besides, two convolutional layers with a simple connection structure are not deep enough to mine the intrinsic radio characteristics for modulation pattern discrimination. Consequently, beside high complexity, BCI-AMR reported low accuracy with higher-order modulations. Whereas, MCNet with an ingenious-designed architecture has obtained a remarkable accuracy while keeping a low network complexity. Concerning deep architecture design, some common instructions can be useful as follows:

- Complexity reduction: the utilization of max/average pooling layers, global average pooling layers, grouped convolutional layers.
- Accuracy improvement: the deployment of skip connection (or residual connection), dense connection, and attention connection besides batch normalization and dropout layers to generalize model.
- Advanced architectures: different fusion mechanism among data-based, feature-based and decision-based approaches, some hybrid networks incorporating FFNN, RNN/LSTM, and CNN are also recommended to delve deeply into performance balance.

Some compression and acceleration techniques [147] can be applied to speed up the training (or learning) and prediction (or inferencing) stages, which are generally categorized into four groups: network pruning, transfer learning, low-rank factorization, and knowledge distillation [148]. Another solution to overcome the high complexity of deep models is the decentralized learning, where a single model is trained on multiple servers with local datasets distributed identically. Recently, federated learning (a.k.a. collaborative learning) [118], [119], [149] has been introduced to train a machine learning algorithm (e.g., deep networks) across multiple decentralized edge devices on heterogeneous datasets (i.e., without exchanging training data for security and privacy). The core principle of this learning framework includes training local models on local datasets individually and sharing learned parameters to generate a global model [150]. Consequently, federated learning for deep networks is a promising solution to obtain multiple objectives: data security and privacy, high accuracy with low latency, and less power consumption for edge devices.

V. CONCLUSION

In this work, we presented a survey of DL-based approaches for accurate modulation classification in wireless communications. At first, some traditional LB and FB approaches were reviewed briefly, where we pointed out their essential drawbacks in terms of reliability and applicability. Then, the fundamental concepts of different primary DL architectures, such as FFNN, RNN/LSTM, and CNN, were elaborated in an extensive manner. Afterwards, the comprehensive reviews and comparisons of state-of-the-art DL-based AMC approaches were provided, along with the insightful analysis of architecture design and mode learning. Finally, we posed a number of key challenges and offered some future research directions on the topic of DL empowered modulation classification. Along with interest, this paper can supply the basic background and advanced knowledge of modulation identification towards other signal processing applications to academic and industrial audiences.

REFERENCES

[1] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, “Survey on 6G frontiers: Trends, applications, requirements, technologies and future research,” IEEE Open J. Commun. Soc., vol. 2, pp. 836–886, 2021.
[2] T. Huynh-The, C.-H. Hua, V.-S. Doan, Q.-V. Pham, and D.-S. Kim, “Accurate deep CNN-based waveform recognition for intelligent radar systems,” IEEE Commun. Lett., vol. 25, no. 9, pp. 2938–2942, Sep. 2021.
[3] T. Huynh-The, V.-S. Doan, C.-H. Hua, Q.-V. Pham, T.-V. Nguyen, and D.-S. Kim, “Accurate LPI radar waveform recognition with CWD-TFA for deep convolutional network,” IEEE Wireless Commun. Lett., vol. 10, no. 8, pp. 1638–1642, Aug. 2021.
[4] V.-S. Doan, T. Huynh-The, and D.-S. Kim, “Underwater acoustic target classification based on dense convolutional neural network,” IEEE Geosci. Remote Sens. Lett., early access, Oct. 19, 2020, doi: 10.1109/LGRS.2020.3029584.
[5] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” IEEE Trans. Cogn. Commun. Netw., vol. 3, no. 4, pp. 563–575, Dec. 2017.
[6] A. Swami and B. M. Sadler, “Hierarchical digital modulation classification using cumulants,” IEEE Trans. Commun., vol. 48, no. 3, pp. 416–429, Mar. 2000.
[7] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, “Survey of automatic modulation classification techniques: Classical approaches and new trends,” IET Commun., vol. 1, no. 2, p. 137, Apr. 2007.
[8] Z. Zhu and A. K. Nandi, Automatic Modulation Classification: Principles, Algorithms and Applications. River Street Hoboken, NJ, USA: Wiley, 2015.
[9] W. Su, J. L. Xu, and M. Zhou, “Real-time modulation classification based on maximum likelihood,” IEEE Commun. Lett., vol. 12, no. 11, pp. 801–803, Dec. 2008.
[10] J. L. Xu, W. Su, and M. Zhou, “Likelihood-ratio approaches to automatic modulation classification,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 41, no. 4, pp. 455–469, Jul. 2011.
[11] O. Ozdemir, R. Li, and P. K. Varshney, “Hybrid maximum likelihood modulation classification using multiple radios,” IEEE Commun. Lett., vol. 17, no. 10, pp. 1889–1892, Oct. 2013.
[12] S. Majhi, R. Gupta, W. Xiang, and S. Glisic, “Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals,” IEEE Trans. Veh. Technol., vol. 66, no. 12, pp. 11057–11069, Dec. 2017.
[13] V. D. Orlic and M. L. Dukic, “Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions,” IEEE Commun. Lett., vol. 13, no. 12, pp. 917–919, Dec. 2009.
[14] B. Ramkumar, “Automatic modulation classification for cognitive radios using cyclic feature detection,” IEEE Circuits Syst. Mag., vol. 9, no. 2, pp. 27–45, Jun. 2009.
[15] W. Su, “Feature space analysis of modulation classification using very high-order statistics,” IEEE Commun. Lett., vol. 17, no. 9, pp. 1688–1691, Sep. 2013.
[16] X. Ning, P. Duan, W. Li, and S. Zhang, “Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer,” IEEE Signal Process. Lett., vol. 27, pp. 1944–1948, 2020.
[17] S. Qi, X. Ning, G. Yang, L. Zhang, P. Long, W. Cai, and W. Li, “Review of multi-view 3D object recognition methods based on deep learning,” Displays, vol. 69, Sep. 2021, Art. no. 102053.
[18] F. L. Luo, Neural Networks for Signal Intelligence. Piscataway, NJ, USA: IEEE, 2020, pp. 243–264.
[19] G. D’Angelo and F. Palmieri, “Network traffic classification using deep convolutional recurrent autoencoder neural networks for spatial–temporal features extraction,” J. Netw. Comput. Appl., vol. 173, Jan. 2021, Art. no. 102890.
[20] J. Clement, N. Indira, P. Vijayakumar, and R. Nandakumar, “Deep learning-based modulation classification for 5G and beyond wireless systems,” Peer Peer Netw. Appl., vol. 14, pp. 319–332, Jan. 2021.
[21] S. Gamjage and J. Samarabandu, “Deep learning methods in network intrusion detection: A survey and an objective comparison,” J. Netw. Comput. Appl., vol. 169, Nov. 2020, Art. no. 102767.

[22] S. Zhou, Z. Yin, Z. Wu, Y. Chen, N. Zhao, and Z. Yang, “A robust modulation classification method using convolutional neural networks,” EURASIP J. Adv. Signal Process., vol. 2019, no. 1, pp. 1–15, Dec. 2019.

[23] T. Zhang, C. Shuai, and Y. Zhou, “Deep learning for robust automatic modulation recognition method for IoT applications,” IEEE Access, vol. 8, pp. 117689–117697, 2020.

[24] R. Zhou, F. Liu, and C. W. Gravelle, “Deep learning for modulation recognition: A survey with a demonstration,” IEEE Access, vol. 8, pp. 67366–67376, 2020.

[25] Q. Mao, F. Hu, and Q. Hao, “Deep learning for intelligent wireless network: A comprehensive survey,” IEEE Commun. Surveys Tut., vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.

[26] X. Li, F. Dong, S. Zhang, and W. Guo, “A survey on deep learning techniques in wireless signal recognition,” Wireless Commun. Mobile Comput., vol. 2019, pp. 1–12, Feb. 2019.

[27] B. Jidd, K. Hassan, I. Dayoub, W. H. Lim, and M. Mokayed, “Machine learning based automatic modulation recognition for wireless communications: A comprehensive survey,” IEEE Access, vol. 9, pp. 57851–57873, 2021.

[28] Q.-V. Pham, N. T. Nguyen, T. Huynh-The, L. Bao Le, K. Lee, and W.-J. Hwang, “Intelligent radio signal processing: A survey,” IEEE Access, vol. 9, pp. 83818–83850, 2021.

[29] H. Hu, Z. Liu, and J. An, “Mining mobile intelligence for wireless systems: A deep neural network approach,” IEEE Comput. Intelleg., Mag., vol. 15, no. 1, pp. 31–41, Feb. 2020.

[30] W. Luo, J. Lu, X. Li, L. Chen, and K. Liu, “Rethinking motivation of deep neural architectures,” IEEE Circuits Syst. Mag., vol. 20, no. 4, pp. 65–76, Nov. 2020.

[31] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[32] N. E. Barabanov and D. V. Prokhorov, “Stability analysis of discrete-time recurrent neural networks,” IEEE Trans. Neural Netw., vol. 13, no. 2, pp. 292–303, Mar. 2002.

[33] Q. Song, Y. Wu, and Y. C. Soh, “Robust adaptive gradient-descent training algorithm for recurrent neural networks in discrete time domain,” IEEE Trans. Neural Netw., vol. 19, no. 11, pp. 1841–1853, Nov. 2008.

[34] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[35] U. Rauter, R. K. Bhattacharya, N. M. Kooi, M. M. Kandim, B. R. Steunenberg, and J. Schmidhuber, “LSTM: A search space Odyssey,” IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2222–2232, Oct. 2017.

[36] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, “Deep learning with long short-term memory for time series prediction,” IEEE Commun. Mag., vol. 57, no. 6, pp. 114–119, Jun. 2019.

[37] F. A. Gers, “Learning to forget: Continual prediction with LSTM,” in Proc. 9th Int. Conf. Artif. Neural Netw. (ICANN), Edinburgh, U.K., 1999, pp. 2451–2471.

[38] S. N. Tran, A. D. Garcez, T. Weyde, J. Yin, Q. Zhang, and M. Karunanithi, “Sequence classification restricted Boltzmann machines with gated units,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 11, pp. 4806–4815, Nov. 2020.

[39] T. Huynh-The, C.-H. Hua, T.-T. Ngo, and D.-S. Kim, “Image representation of pose-transition feature for 3D skeleton-based action recognition,” Inf. Sci., vol. 514–515, pp. 122–126, Mar. 2020.

[40] C.-H. Hua, T. Huynh-The, S.-H. Bae, and S. Lee, “Cross-attentional bracket-shaped convolutional network for semantic image segmentation,” Inf. Sci., vol. 539, pp. 277–294, Oct. 2020.

[41] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, “Learning 3D spatiotemporal gait feature by convolutional network for person identification,” Neurocomputing, vol. 397, pp. 192–202, Jul. 2020.

[42] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, “Robust image watermarking framework powered by convolutional encoder-decoder network,” in Proc. Digit. Image Comput.: Techn. Appl. (DICTA), Perth, WA, Australia, Dec. 2019, pp. 1–7.

[43] T. Hur, J. Bang, T. Huynh-The, J. Lee, J.-I. Kim, and S. Lee, “Iss2Image: A novel signal-encoding technique for CNN-based human activity recognition,” Sensors, vol. 18, no. 11, pp. 3910, Nov. 2018.

[44] R. C. Gonzalez, “Deep convolutional neural networks [lecture notes],” IEEE Signal Process. Mag., vol. 35, no. 6, pp. 79–87, Nov. 2018.
T. Huynh-The et al.: Automatic Modulation Classification: Deep Architecture Survey
[150] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, “Federated learning in mobile edge networks: A comprehensive survey,” IEEE Commun. Surveys Tuts., vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

THIEN HUYNH-THE (Member, IEEE) received the B.S. degree in electronics and telecommunication engineering from the Ho Chi Minh City University of Technology and Education, Vietnam, in 2011, and the Ph.D. degree in computer science and engineering from Kyung Hee University (KHU), South Korea, in 2018.

From March 2018 to August 2018, he was at KHU. He is currently a Postdoctoral Research Fellow with the ICT Convergence Research Center, Kumoh National Institute of Technology, South Korea. His current research interests include radio signal processing, digital image processing, computer vision, wireless communications, and deep learning. He is awarded with the Superior Thesis Prize by KHU.

QUOC-VIET PHAM (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering from the Hanoi University of Science and Technology, Vietnam, in 2013, and the Ph.D. degree in telecommunications engineering from Inje University, South Korea, in 2017.

From September 2017 to December 2019, he was with Kyung Hee University, Changwon National University, and Inje University, on various academic positions. He is currently a Research Professor with Pusan National University, South Korea. He has been granted the Korea NRF Funding for Outstanding Young Researchers for the term 2019–2023. His research interests include convex optimization, game theory, and machine learning to analyze and optimize edge/cloud computing systems and future wireless systems. He received the Best Ph.D. Dissertation Award in Engineering from Inje University, in 2017. He received the top Reviewer Award from the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, in 2020. He is an Editor Journal of Network and Computer Applications (Elsevier) and a Lead Guest Editor of the IEEE INTERNET OF THINGS JOURNAL.

TOAN-VAN NGUYEN (Member, IEEE) received the B.S. degree in electronics and telecommunications engineering and the M.S. degree in electronics engineering from the HCMC University of Technology and Education, Vietnam, in 2011 and 2014, respectively, and the Ph.D. degree in electronics and computer engineering from Hongik University, South Korea, in 2021. He is currently a Postdoctoral Researcher with the Electrical and Computer Engineering Department, Utah State University, Logan, UT, USA. His current research interests include mathematical modeling of 5G networks and machine learning for wireless communications.

THANH THI NGUYEN received the Ph.D. degree in mathematics and statistics from Monash University, Australia, in 2013. He was a Visiting Scholar with the Computer Science Department, Stanford University, CA, USA, in 2015, and the Edge Computing Laboratory, Harvard University, MA, USA, in 2019. He is currently a Senior Lecturer in AI at the School of Information Technology, Deakin University, VIC, Australia. He has expertise in artificial intelligence, deep learning, reinforcement learning, big data, neuroscience, signal and image processing, cybersecurity, the IoT, and multi-agent systems. He received the Alfred Deakin Postdoctoral Research Fellowship, in 2016, the European–Pacific Partnership for ICT Expert Exchange Program Award from European Commission, in 2018, and the Australia–India Strategic Research Fund Early- and Mid-Career Fellowship awarded by the Australian Academy of Science, in 2020.

RUKHSANA RUBY (Member, IEEE) received the master’s degree from the University of Victoria, Victoria, Canada, in 2009, and the Ph.D. degree from The University of British Columbia, BC, Canada, in 2015. From the broader aspect, her resource interests include the management and optimization of next generation wireless networks. She has authored nearly 70 technical articles of well-recognized journals and conferences. Besides, she has also been serving as a technical program committee member in various reputed conferences. She was a recipient of several awards or honors, notable among which are Wait-listed for Canadian NSERC Postdoctoral Fellowship, the IEEE Exemplary Certificate from IEEE COMMUNICATIONS LETTERS, in 2018, and IEEE WIRELESS COMMUNICATIONS LETTERS, in 2018; and the Outstanding Reviewer Certificate from Computer Communications (Elsevier), in 2017. She has served as the Lead Guest Editor for the Special Issue on NOMA techniques under EURASIP JWCN, in 2017. She is currently serving as an Associate Editor for EURASIP JWCN.

MING ZENG (Member, IEEE) received the B.E. and master’s degrees from the Beijing University of Posts and Telecommunications, China, in 2013 and 2016, respectively, and the Ph.D. degree in telecommunications engineering from Memorial University, Canada, in 2020. He is currently an Assistant Professor with the Department of Electrical Engineering and Computer Engineering, Laval University, Canada. He has published more than 45 articles and conferences in first-tier IEEE journals and proceedings and his work has been cited over 1150 times per Google Scholar. His research interests include resource allocation for beyond 5G systems and machine learning empowered optical communications. He serves as an Associate Editor for the IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY.

DONG-SEONG KIM (Senior Member, IEEE) received the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, Republic of Korea, in 2003. From 1994 to 2003, he worked as a full-time Researcher at the ERC-ACI, Seoul National University, Seoul. From March 2003 to February 2005, he worked as a Postdoctoral Researcher with the Wireless Network Laboratory, School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, USA. From 2007 to 2009, he was a Visiting Professor with the Department of Computer Science, University of California at Davis, Davis, CA, USA. He is currently the Director with the ICT Convergence Research Center (Grand ICT Program), Kit Convergence Research Institute supported by the Korean Government with the Kumoh National Institute of Technology. His current main research interests include the real-time IoT, industrial wireless control networks, networked embedded systems, and fieldbus. He is a Senior Member of ACM.

* * *