2
Signal Model and Literature Survey

2.1 INTRODUCTION

All the modulation classification algorithms are developed and tested based on the particular channel scenario and signal models are needed to be defined before algorithm development. Signal statistics change while transmission through the channel and decision rules of likelihood and feature-based methods depends on the distribution of the channel model. Even though most of the deep learning methods do not require the signal model to be known for classification but it helps to design an optimum network for better classification performance.

This chapter defines all the signal models used in this thesis for the development of algorithms and testing. Three-channel models, AWGN, flat fading, and Gaussian mixer model (GMM) are considered in this research to test the developed method performance. The literature review of the existing modulation classifiers like likelihood-based, feature-based as well as some recent developments in machine learning and deep learning-based classifiers are included in this chapter to understand the status of the developed classifiers.

2.1.1 Signal Model in AWGN channel

The AWGN is the most commonly used noise model in communication. The source of this noise is the thermal vibration present in the conductors. AWGN is additive noise and its effect comes due to the direct addition with the baseband signal. Mostly, white noise is generated due to the receiver hardware and present in both wired and wireless communication systems. The spectral density of the AWGN is constant for all the frequency range for a fixed system.

After estimation and correction of all impairments, the received baseband signal is downsampled with the symbol rate to extract the complex symbols. In presence of AWGN, the symbols can be represented by

$$r[n] = a[n] + \eta[n] \tag{2.1}$$

Here $a[n]$ is the n^{th} constellation point with $a_I[n]$ be the in-phase component and $a_Q[n]$ be the quadrature component of that symbol and $\eta[n] = \eta_I[n] + j\eta_Q[n]$ is complex Gaussian noise with the probability density function (PDF) given by

$$f_\eta(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{|x|^2}{2\sqrt{|\Sigma|}}\right) \tag{2.2}$$

Here Σ is the covariance matrix of AWGN, $|\Sigma|$ is the determinant value of AWGN covariance matrix, and $|x|$ is the amplitude absolute value of complex noise samples. The covariance matrix of AWGN can be represented as

$$\Sigma = \begin{bmatrix} \sigma_{\eta_I}^2 & \rho\sigma_{\eta_I}\sigma_{\eta_Q} \\ \rho\sigma_{\eta_I}\sigma_{\eta_Q} & \sigma_{\eta_Q}^2 \end{bmatrix} \tag{2.3}$$

Here variance of I and Q-phase component of complex AWGN is given by $\sigma_{\eta_I}^2$ and $\sigma_{\eta_Q}^2$, respectively. ρ is correlation between I-phase and Q-phase component. In some of the applications I and Q components are required separately, so to get the PDF of both the components is essential. Both the components are independent and identically distributed (IID), hence $\rho = 0$ and updated Σ can be given by

$$\Sigma = \begin{bmatrix} \sigma_{\eta_I}^2 & 0 \\ 0 & \sigma_{\eta_Q}^2 \end{bmatrix} \tag{2.4}$$

The I and Q components are independent and their PDF can be represented by the expression given in Equation (2.5).

$$f_{\eta_I}(x) = f_{\eta_Q}(x) = \frac{1}{\sqrt{2\pi}\sigma_\eta} \exp\left(-\frac{|x|^2}{2\sigma^2}\right) \tag{2.5}$$

2.1.2 Signal in Fading channel

In a practical channel environment, a signal gets a significant effect on its amplitude, phase, and frequency and corresponding impairments are known as attenuation ($\hat{\alpha}$), phase offset ($\Delta\phi$), and frequency offset (Δf). Generally, two types of fading channels are considered, one is a slow fading channel which is caused due to the shadowing effect of the environment. The coherence time of the slow fading channel is much larger than symbol time and the channel can be represented by constant attenuation and phase offset concerning time for a certain number of symbols. The signal model of the slow fading channel can be given as

$$r(t) = \hat{\alpha} e^{j\Delta\phi} s(t) + \eta(t) \tag{2.6}$$

The second type of channel is fast fading, in this channel condition signal reflects from objects in the surrounding of the receiver antenna and all the signal copies get added in different phases which causes a drastic change in the amplitude and phase. This channel has a small coherence time as compared to the symbol time and amplitude and phase changes with time. The signal model for the fast fading channel can be given as

$$r(t) = \alpha(t)s(t) + \eta(t) \tag{2.7}$$

Here $\alpha(t)$ has included time varying attenuation and phase offset as per the expression $\alpha(t) = \hat{\alpha}(t)e^{j\Delta\phi(t)}$. $\hat{\alpha}(t)$ and $\Delta\phi(t)$ are assumed to be Gaussian distributed with $\hat{\alpha}_0$, $\Delta\phi$ mean and $\sigma_{\hat{\alpha}}$, σ_ϕ variance, respectively. Apart from above mentioned impairments, frequency offset imposes significant effect on the signal traversing in the channel. For given carrier frequency (f_c), if the receiver is moving with velocity v, there will be $\Delta f_c = f_c v/c$ frequency shift in carrier frequency due to the Doppler effect. This frequency offset causes rotation of the symbols and classification gets difficult. With inclusion of all impairments, the signal model is expressed by

$$r(t) = \hat{\alpha}(t)e^{j(2\pi\Delta ft+\Delta\phi)}s(t)) + \eta(t) \tag{2.8}$$

2.1.3 Signal model for Gaussian Mixture Model

Middleton has given expressions for a series of non-Gaussian noises to approximate the noise generated by electromagnetic activities. One of the noise models is the class A model which represents non-Gaussian noises having a bandwidth less than the bandwidth of the signal. Vastola has approximated the class A noise model by using a mixture of Gaussian noises which is being used frequently because of less computation complexity. The expression of the Gaussian mixer model (GMM) is given by

$$f_\eta(p) = \sum_{n=1}^{N} \frac{\lambda_n}{2\pi\sigma^2} e^{-\frac{|p|^2}{2\sigma^2{}_n}} \tag{2.9}$$

Here N is the number of Gaussian noise components, λ_n is probability to choose n^{th} noise component, and σ_n^2 represents variance of n^{th} noise component. It is assumed that all the noises have zero mean.

2.2 REVIEW OF EXISTING CLASSIFIERS

The modulation classifiers are broadly categorized into two parts: Likelihood-based (LB) and Feature-based (FB). LB methods have been interesting as these convert the classification problem into a hypothesis testing problem. It is seen in the literature that these methods provide the optimum classification accuracy while the channel statistics are known to the receiver. Because multiple hypotheses get tested, the computation complexity is high and in the case of unknown parameters, complexity becomes even higher. On the other hand, FB methods use features extracted from raw signal or preprocessed signal followed by a suitable classifier to provide the classification. These are not optimal classifiers but have low computation complexity which makes them advantageous for real-time applications.

2.2.1 Likelihood-based Classifiers

The likelihood-based classification method is executed into two steps: likelihood calculation and likelihood comparison. In the first step, the value of the likelihood function is calculated which indicates that how closely the unknown signal symbols are matching with one particular modulation. In the second step, the likelihood function comparison is done for the final classification decision. In the case of FB classification methods the CSI at the receiver needs to be known otherwise classification performance reduces drastically but for LB methods, in case of unknown CSI conditions, the likelihood function is updated according to the available parameters. The likelihood function has been modified in the maximum likelihood classifier, maximum likelihood ratio test, generalized likelihood ratio test, and hybrid likelihood ratio test to resolve the issue of unknown CSI. All of these methods are discussed in this section.

Maximum likelihood classifier

The maximum likelihood classifier is optimum. It computes the probability of received signal samples belongs to the model of a particular modulation scheme with the given channel parameters. Modulation classification decision is given for which the likelihood function takes the maximum value.

Assuming that the received signal sample $r[n]$ belongs to modulation $\mathfrak{M}$, likelihood function value or probability of that point to be observed with the given AWGN model is

$$\mathfrak{L}(r[n]|\mathfrak{M},\sigma) = \mathscr{P}(r[n]|\mathfrak{M},\sigma) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{|r[n]-a_m|^2}{2\sigma^2}\right) \tag{2.10}$$

Here a_m is a complex symbol point that belongs to a particular modulation scheme, M is the number of expected symbol points from selected modulation and σ is the variance of AWGN. For total K received symbols, the joint likelihood function is given by

$$\mathfrak{L}(r[n]|\mathfrak{M},\sigma) = \prod_{k=1}^{K}\frac{1}{M}\sum_{m=1}^{M}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{|r[n]-a_m|^2}{2\sigma^2}\right) \tag{2.11}$$

Sometimes, for analytical simplicity, natural logarithm is applied to get log-likelihood function, i.e.,

$$log\mathfrak{L}(r[n]|\mathfrak{M},\sigma) = log\left(\prod_{k=1}^{K}\frac{1}{M}\sum_{m=1}^{M}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{|r[n]-a_m|^2}{2\sigma^2}\right)\right) \tag{2.12}$$

$$log\mathfrak{L}(r[n]|\mathfrak{M},\sigma) = \sum_{k=1}^{K}log\left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{2\pi\sigma^2}\exp\left(-\frac{|r[n]-a_m|^2}{2\sigma^2}\right)\right) \tag{2.13}$$

After calculating the likelihood function with the given channel parameters, values for all modulation schemes are compared and the decision is given in favor of modulation having maximum likelihood function value.

Likelihood Ratio Test

To calculate the likelihood function, channel parameters have to be known. In the case of unknown channel parameters, classification accuracy decreases. This drawback is overcome by Polydoros and Kim in 1990. They have developed an average likelihood ratio test (ALRT) which considers unknown parameters as a random variable with a certain PDF. In this method, they have considered the unknown parameter as a random variable with a certain PDF. The likelihood function value is obtained by integrating the likelihood function with the joint PDF of parameters for an applicable range of one or more unknown parameters. Considering that unknown channel parameters set Φ consisting of phase offset ($\Delta\phi$) and channel gain ($\hat{\alpha}$), the ALRT likelihood function can be expressed by

$$\mathfrak{L}_{ALRT}(r) = \int_{\Phi}\mathfrak{L}(r|\mathfrak{M})f(\Phi|H)d\Phi \tag{2.14}$$

$$= \int_{\Phi} \prod_{k=1}^{K} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|r[n] - \hat{\alpha} e^{-j\Delta\phi} a_m|^2}{2\sigma^2}\right) f(\Phi|H) d\Phi \tag{2.15}$$

Here $\mathfrak{L}(r|\mathfrak{M})$ is likelihood value with available channel parameters Φ and $f(\Phi|H)$ is joint PDF of parameters Φ for modulation hypothesis H. It can be seen that ALRT classifiers are more complex in comparison to maximum likelihood due to the involvement of integration and complexity increases with more unknown parameters. Also, PDF of unknown channel parameters is required to be known accurately otherwise performance becomes suboptimal. Hence, estimation of parameters is an important part which leads it to more complex and inaccurate. To resolve the complexity issue Panagiotou, Anastasopoulos, and Polydoros have proposed a generalized likelihood ratio test (GLRT) which maximizes likelihood function for all possible values of unknown parameters. The likelihood of GLRT for the unknown parameter set Φ is given with the expression

$$\mathfrak{L}_{GLRT}(r) = max_{\Phi} \mathfrak{L}(r|\hat{\alpha}, \sigma, \Delta\phi) \tag{2.16}$$

$$= max_{\Phi} \prod_{k=1}^{K} \frac{1}{M} \sum_{m=1}^{M} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{|r[n] - \hat{\alpha} e^{-j\Delta\phi} a_m|^2}{2\sigma^2}\right) \tag{2.17}$$

Complexity in GLRT is decreased but it becomes biased to the higher-order modulation in the case of nested modulation schemes because of common symbol points. To combat this problem Panagiotou has developed a hybrid likelihood ratio test (HLRT) by the combination of ALRT and GLRT, where the likelihood is averaged over all symbol points, and then the resulting function is maximized for all unknown parameters.

2.2.2 Feature Based Classification

Likelihood-based classifiers provide optimum classification efficacy but with high computation complexity, because of that, they cannot be used for real-time applications. Feature-based classification methods provide near optimum performance with less computation complexity.

FB classification is divided into two stages: feature extraction and classification. In the first stage, some important time-domain or frequency-domain features are extracted and in the second stage, a suitable classifier is employed for modulation classification. In this section, some of the spectral-based features are discussed which have been used by Nandi and Azzouz for the classification of digital and analog modulation schemes using decision tree classifiers [Nandi and Azzouz, 1995; Azzouz and Nandi, 1995]. At the same time, higher-order statistics and cyclostationary based features are exploited for the classification of various modulation schemes of different orders.

Time-domain and Frequency-domain Features

The modulation classification of digital and analog modulation schemes have been done using spectral features of basic aspects of the signals viz. amplitude, phase, and frequency. A decision tree-based classifier is employed to implement the flow of the classification process. The first feature is γ_{max}, the maximum value of the spectral power density of centered instantaneous and normalized amplitude of the received signal, which is given as

$$\gamma_{max} = \frac{max|DFT(A_{cn})|^2}{N} \tag{2.18}$$

Here N represents number of received sample points and DFT stands for Discrete Fourier Transform. A_{cn} is the centered instantaneous and normalized amplitude, which can be given by

$$A_{cn}[n] = A_n[n] - 1 \tag{2.19}$$

where, $A_n[n] = A_n/\mu_A$ and μ_A is mean value of the instantaneous amplitude. The second feature, σ_{ap} is the standard deviation of the absolute value of the non-linear component of the instantaneous phase.

$$\sigma_{ap} = \sqrt{\frac{1}{N_t}\left(\sum_{A_n[n]>A_{th}} \phi_{nl}^2[n]\right) - \left(\frac{1}{N_t}\sum_{A_n[n]>A_{th}} |\phi_{nl}[n]|\right)^2} \tag{2.20}$$

where N_t is number of samples which satisfies the condition $A_n[n] > A_{th}$. This condition is applied to remove the samples which are more prone to noise. The third feature, σ_{dp}, is the standard deviation of the non-linear component of the direct instantaneous phase.

$$\sigma_{dp} = \sqrt{\frac{1}{N_t}\left(\sum_{A_n[n]>A_{th}} \phi_{nl}^2[n]\right) - \left(\frac{1}{N_t}\sum_{A_n[n]>A_{th}} \phi_{nl}[n]\right)^2} \tag{2.21}$$

The fourth feature is the spectrum symmetry around the carrier frequency and denoted by $\mathbb{S}$.

$$\mathbb{S} = \frac{\mathbb{S}_L - \mathbb{S}_U}{\mathbb{S}_L + \mathbb{S}_U} \tag{2.22}$$

$$\mathbb{S}_L = \sum_{n=1}^{f_{cn}} |Y_c[n]|^2 \tag{2.23}$$

$$\mathbb{S}_U = \sum_{n=1}^{f_{cn}} |Y_c[n + f_{cn} + 1]|^2 \tag{2.24}$$

Y_c is Fourier Transform of $y_c[n]$, $f_{cn} + 1$ is sample index corresponding to carrier frequency f_c, f_s is sampling frequency, and $f_{cn} = (f_c N/f_s) - 1$.

The fifth feature σ_{aa} is the standard deviation of the absolute value of the normalized and centered instantaneous amplitude and the sixth feature σ_{af} is the standard deviation of the absolute value of the normalized and centered instantaneous frequency.

$$\sigma_{aa} = \sqrt{\frac{1}{N_c}\left(\sum_{n=1}^{N} A_{cn}^2[n]\right) - \left(\frac{1}{N_c}\sum_{n=1}^{N} |A_{cn}[n]|\right)^2} \tag{2.25}$$

$$\sigma_{af} = \sqrt{\frac{1}{N_t}\left(\sum_{A_n[n]>A_{th}} f_n^2[n]\right) - \left(\frac{1}{N_t}\sum_{A_n[n]>A_{th}} |f_n[n]|\right)^2} \tag{2.26}$$

where $f_n[n]$ is the centred instantaneous frequency normalized by the sampling frequency and $f_m[n]$ is centred instantaneous frequency with the frequency mean μ_f.

$$f_n[n] = f_m[n]/f_s \tag{2.27}$$

$$f_m[n] = f[n] - \mu_f \tag{2.28}$$

$$\mu_f = \frac{1}{N}\sum_{n=1}^{N} f[n] \tag{2.29}$$

The seventh feature, σ_a is the standard deviation of normalized and centered instantaneous amplitude.

$$\sigma_{a} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n]>A_{th}} a_{cn}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n]>A_{th}} a_{cn}[n]\right)^2} \tag{2.30}$$

The eight and ninth features μ_{42}^a, μ_{42}^f are kurtosis of the normalized and centered instantaneous amplitude and frequency.

$$\mu_{42}^a = \frac{E\{A_{cn}^4[n]\}}{\{E\{A_{cn}^2[n]\}\}^2} \tag{2.31}$$

$$\mu_{42}^f = \frac{E\{f_N^4[n]\}}{\{E\{f_N^2[n]\}\}^2} \tag{2.32}$$

Azzouz and Nandi have designed a decision tree-based classifier which consists of the features as an input and a sequence of conditions is employed for classification between VSB, M-ASK, AM, M-FSK, FM, M-QAM, M-PSK, DSB and, SSB modulation schemes as depicted in Figure 2.1.

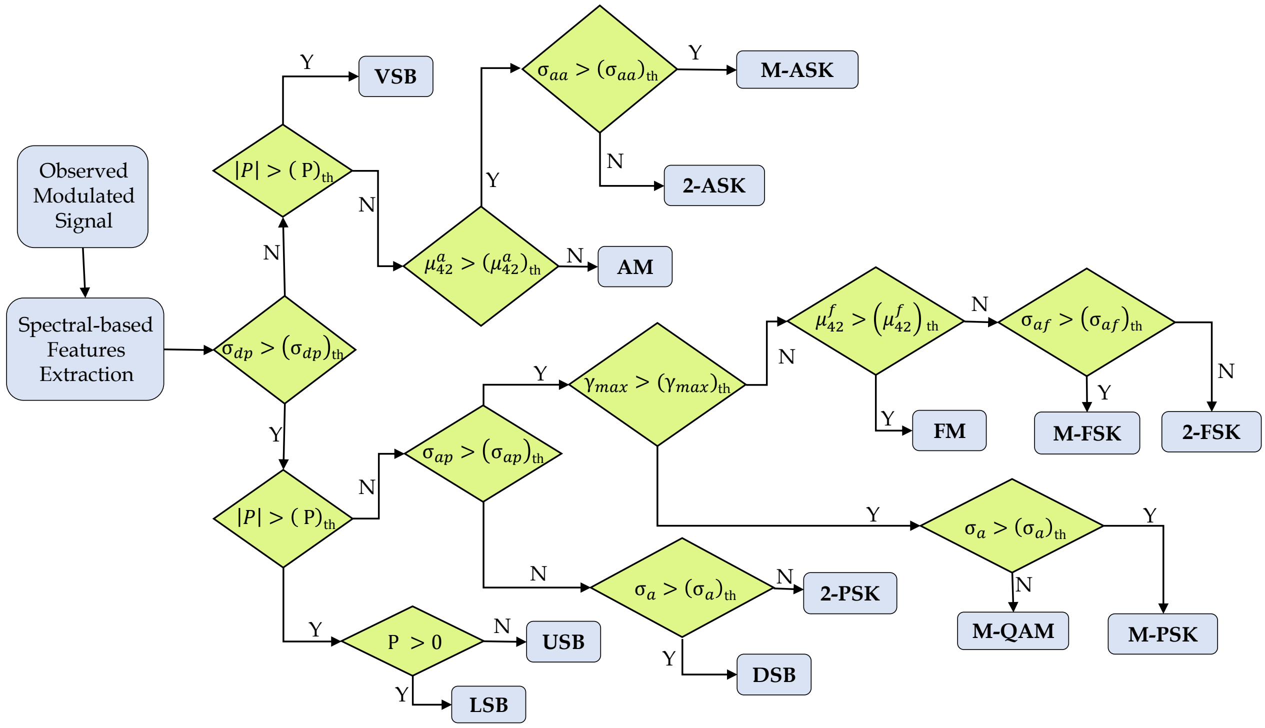

Figure 2.1 : Spectral feature based classification strategy

High-order Statistical features

In 1986, Hipp has employed the third-order moment of the amplitude of the signal as a classification feature [Hipp, 1986]. Soliman and Hsue have extended this work by investigating higher-order moments of signal phase for M-PSK classification [Soliman and Hsue, 1992]. It has been observed that k^{th} moment of signal phase is increasing function of PSK modulation order. The higher-order PSK modulation schemes have higher moment values and can be classified using appropriate threshold values. The k^{th} moment of signal phase can be represented as

$$\mu_k[n] = \frac{1}{N}\sum_{n=1}^{N}\phi^k[n] \tag{2.33}$$

where $\phi[n]$ stands for phase of n^{th} sample point. Also, Swami and Sadler have investigated different order cumulants for the classification of M-QAM, M-PSK and, M-PAM modulations Swami and Sadler [2000]. The expression for second and fourth-order cumulants can be given as

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} r^2[n] \tag{2.34}$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} |r[n]|^2 \tag{2.35}$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} r^4[n] - 3\hat{C}_{20} \tag{2.36}$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} r^3[n]r^*[n] - 3\hat{C}_{20}\hat{C}_{21} \tag{2.37}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |r[n]|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2 \tag{2.38}$$

Cyclostationary based features

In 1994, Gardner has investigated the periodic characteristics of the cyclostationay signal for classification. Later on, Spooner and Gardner have found diverse spectrum appearance for different modulation [Gardner and Spooner, 1988]. In 2009, Ramkumar has summarized the cyclic feature-based modulation classification [Ramkumar, 2009]. The cyclic autocorrelation of a signal $r(t)$ can be expressed as

$$R_r^{\alpha}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} r\left(t+\frac{\tau}{2}\right) r\left(t-\frac{\tau}{2}\right) e^{-j2\pi\alpha t}dt \tag{2.39}$$

where cyclic frequency $\alpha \neq 0$. The Spectral Correlation Function (SCF) of $r(t)$ using Wiener-Khintchine theorem is given by

$$S_r(f) = \int_{-\infty}^{\infty} R_r^{\alpha}(\tau)e^{-j2\pi f\tau}d\tau \tag{2.40}$$

2.2.3 Machine Learning based classifiers

In the previous section, different features were explained. The classification of modulation is achieved by feeding these features to the hierarchical decision trees. The investigation for optimum decision thresholds is complex to implement. Therefore, machine learning (ML) has been used for modulation classification in the literature. Machine learning is useful to get the optimum threshold as well as it reduces the feature space by removing the features which are not contributing towards classification. ML classifiers are divided into two categories: supervised and unsupervised. In supervised classifiers, features of the signal are extracted and given to the classifier for training to identify the optimum threshold or parameters. Some of the supervised ML classifier are Artificial Neural Network (ANN), support vector machine (SVM) [Park *et al.*, 2008; Hazza *et al.*, 2013; Han Gang *et al.*, 2004; Teng *et al.*, 2008; Hu *et al.*, 2008; Mustafa and Doroslovacki, 2004], k-nearest neighbor (k-NN) [Aslam *et al.*, 2012, 2010], Naive Bayes classifier [Wong *et al.*, 2008], and logistic regression etc.

Initially in 1998, Nandi and Azzouz have used the ANN for modulation classification of analog and digital signals using different spectral-based features [Nandi and Azzouz, 1998]. The classification process is done in two stages. In the first stage, ANN is trained with spectral features to classify all considered analog and digital modulation schemes except the orders of ASK and

FSK. In the second stage, two ANN are trained to classify between the orders of ASK and FSK. ANN has been used with the other features like cyclic spectral features [Fehske *et al.*, 2005; Qian and Zhu, 2010], statistical signal characteristics [Hossen *et al.*, 2007; Wong and Nandi, 2004], and combination of features [WON, 2004].

Recently, Deep Learning (DL) is being used in wide areas of research like medical image processing [Kermany *et al.*, 2018; Shin *et al.*, 2016; Anthimopoulos *et al.*, 2016], Natural Language Processing (NLP) [Chen *et al.*, 2018], market price forecasting [Fischer and Krauss, 2018] and image classification [Ranjan *et al.*, 2017; Sun *et al.*, 2019; Wang *et al.*, 2019]. On the footnote of image processing, DL has also been introduced in modulation classification [Ali *et al.*, 2017; Peng *et al.*, 2018; O?Shea *et al.*, 2018; Meng *et al.*, 2018; Hu *et al.*, 2019; Wang *et al.*, 2020b; Peng *et al.*, 2019; Mendis *et al.*, 2016; Hauser *et al.*, 2017; Liu *et al.*, 2017; Kim *et al.*, 2016; Peng *et al.*, 2017; Karra *et al.*, 2017; Kim *et al.*, 2016]. DL networks extract significant features at its own without any manual intervention and provides improved classification efficacy. In [Ali *et al.*, 2017], the proposed DL network is trained with three different data viz. complex baseband symbols, estimated centroids using fuzzy C-means algorithm, and high-order cumulants to show the superiority of the deep network over shallow one. In [Peng *et al.*, 2018], the constellation data is converted into a color image using gridlike topology to classify modulation schemes using AlexNet and GoogLeNet CNN-based DL networks. In [O?Shea *et al.*, 2018], authors have considered the higher-order moments and cumulants as an input to the DL model. To show the robustness of the method, authors have considered channel effects like carrier frequency offset, symbol time offset, and multipath fading. It has been seen that the DL algorithms are self-sufficient to extract the significant features from the baseband complex symbol sequence [Meng *et al.*, 2018; Hu *et al.*, 2019]. In [Wang *et al.*, 2020b], authors have developed neuron pruning technology to reduce the model size for fast computation with slight performance loss.

Some of the unsupervised clustering-based methods have also been investigated for modulation classification. These methods do not require group labels for the training of the classification model but use a constellation structure. In [Weber *et al.*, 2015], the K-means algorithm is employed to extract the features used for classification. In [Azarmanesh and Bilén, 2013], a combination of the 'K-centre' and 'K-means' algorithms have been used and the standard deviation of the derived prototypes are used for the modulation classification. In [Norouzi *et al.*, 2016], OFDM and single-carrier signals are distinguished using amplitude statistics of the received signal as features and centroids of I and Q components are estimated using clustering algorithm which has been used for classification. Apart from these other clustering algorithms like hierarchical clustering [Swami and Sadler, 2000], subtractive clustering, and fuzzy c means clustering [Ahmadi, 2010] is also used for feature generation from constellation structure.

…