2
Signal Model and Literature Survey

2.1 SIGNAL MODEL

To start with any AMC algorithm, the signal model has to be defined to design the strategy for the classification of modulation. Signal models are defined according to different real-world scenarios to test the performance of an algorithm in computer-aided simulations. Signal models are also used to calculate the optimized thresholds that are further used in a real channel for efficiently taking decision.

This section aims to define some models used in designing and testing the AMC algorithms for the rest of the thesis. Three-channel models i.e., Additive white Gaussian noise (AWGN), fading channel, and non-Gaussian models are used. There are different probability distribution functions available to approximate noise, but throughout the thesis, AWGN, slow and flat fading, and Gaussian Mixture Model (GMM) is considered for performance evaluation.

2.1.1 AWGN Signal Model

AWGN is a basic noise model in which impairment comes with the linear addition of white noise. White noise is characterized by constant spectral density. The amplitude of the noise follows Normal distribution. AWGN is widely used in wired and wireless communications to characterize the random processes that occur in nature.

To extract the symbol points, the passband signal is downconverted and passes through a matched filter. The complex baseband signal is further downsampled to one sample per symbol. Symbols corrupted by Gaussian noise are represented by:

$$d[n] = s[n] + w[n] \tag{2.1}$$

Here, $s[n] = s_I[n] + js_Q[n]$, i.e., n^{th} complex symbol with in-phase part represented as $s_I[n]$, and quadrature component is represented as $s_Q[n]$. $w[n] = w_I[n] + jw_Q[n]$, i.e., complex Gaussian noise with real and imaginary part following Gaussian distribution with zero mean and σ^2 variance. Probability density function (PDF) of w_I and w_Q are given by

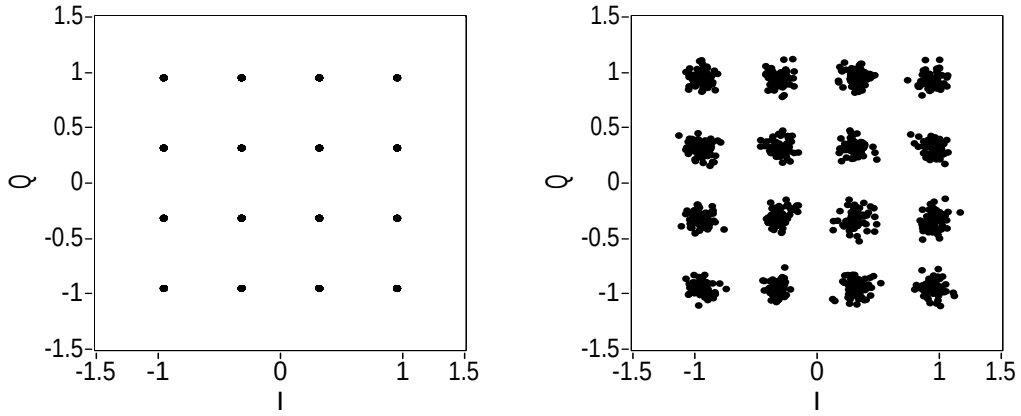
$$f_{w_I}(g) = f_{w_Q}(g) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(g)^2}{2\sigma^2}} \tag{2.2}$$

PDF of the complex normal distribution with Σ as covariance matrix, $|g|$ is the absolute value of complex value g, and $|\Sigma|$ is determinant of Σ is given by

$$f_w(g) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{(|g|)^2}{2\sqrt{|\Sigma|}}} \tag{2.3}$$

Covariance matrix is given by:

$$\Sigma = \begin{bmatrix} \sigma_{w_I}^2 & \rho\sigma_{w_I}\sigma_{w_Q} \\ \rho\sigma_{w_I}\sigma_{w_Q} & \sigma_{w_Q}^2 \end{bmatrix} \tag{2.4}$$



(a) 16-QAM ideal constellation signature.

(b) 16-QAM constellation signature in AWGN channel with SNR of 20dB.

Figure 2.1 : Constellation structure in AWGN noise.

Here $\sigma^2_{w_I}$ and $\sigma^2_{w_Q}$ are the variance of in-phase and quadrature-phase components of complex noise. ρ is the correlation between the in-phase and quadrature component. Both components are independent and identically distributed; hence $\rho = 0$ and Σ is represented by Equation (2.5). Figure 2.1 shows the effect of AWGN noise on symbols of 16-QAM.

$$\Sigma = \begin{bmatrix} \sigma^2_{w_I} & 0 \\ 0 & \sigma^2_{w_Q} \end{bmatrix} \tag{2.5}$$

2.1.2 Signal in Fading channel

In a wireless scenario, the signal reaches the destination taking multiple paths called multipath propagation or shadowing through large objects affecting which causes random fluctuation in propagating signal called fading. The effect of fading is characterized by attenuation, phase offset, and frequency offset in the signal. Fading can be slow or fast, depending on the delay requirement of any application.

When the delay requirement of any application is larger than the coherence time of the channel, fast fading occurs otherwise, it is slowly fading. In fast fading, attenuation a, phase offset θ_0 changes with time. The model for fast fading is given by:

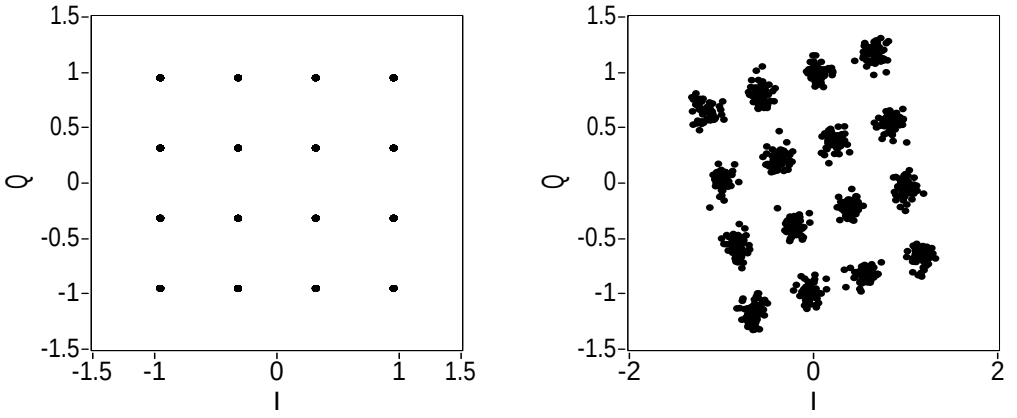
$$y(t) = \alpha(t)x(t) + w(t) \tag{2.6}$$

Here $\alpha(t) = a(t)e^{j\theta_0(t)}$ with $a(t) \sim N(a_0, \sigma_a)$ and $\theta_0(t) \sim N(\theta_0, \sigma_\phi)$. a_0 and θ_0 are mean attenuation and mean phase offset respectively. The variance of attenuation and phase offset are denoted by σ_a and σ_ϕ respectively. Other than attenuation and phase offset, if the receiving antenna is in motion, the Doppler effect comes into the picture. This gives rise to frequency offset, and a signal model with frequency offset f_0 can be expressed as:

$$y(t) = e^{2\pi f_0 t}x(t) + w(t) \tag{2.7}$$

Due to frequency offset, extracted symbol rotates in the IQ plane. The model with all the effects can be expressed as:

$$y(t) = \alpha(t)e^{2\pi f_0 t}x(t) + w(t) \tag{2.8}$$



(a) 16-QAM ideal constellation signature.

(b) 16-QAM constellation signature in fading channel with SNR of 20dB.

Figure 2.2 : Constellation signature of 16-QAM in fading channel.

$$y(t) = a(t)e^{j(2\pi f_0 t + \theta_0(t))}x(t) + w(t) \tag{2.9}$$

For slow fading, a constant channel gain and phase offset are used to model the received signal expressed as:

$$y(t) = ae^{j(2\pi f_0 t + \theta_0)}x(t) + w(t) \tag{2.10}$$

2.1.3 Signal model for Gaussian Mixture Model

The Gaussian mixture model is used to approximate Middleton's class A noise. It is a non-Gaussian noise and comes under the memory-less impulse noise model. The PDF of the GMM model is expressed as

$$f_w(g) = \sum_{n=1}^{K} \frac{\lambda_k}{2\pi\sigma^2} e^{-\frac{|g|^2}{2\sigma_k^2}} \tag{2.11}$$

λ_k is the probability of the noise chosen from the k^{th} component and σ_k^2 is the variance of the k^{th} component. Besides the Middleton Class A model, Bernoulli-Gaussian and Symmetric alpha-stable model also comes under the category of memory-less impulse noise model and Markov-Middleton and Markov-Gaussian are with memory. But out of all these impulse noise models, only GMM is used.

2.2 EXISTING CLASSIFIERS

The field of modulation recognition came into the picture by its use in military scenarios like surveillance, electronic warfare, preparing jamming signals, and recognition of adversary units or jamming the communication link between them. Initially, this task was completed manually by engineers. From the 1980s, numerous classifiers were developed in the literature for various applications. Likelihood-based modulation classifiers [Chung-Yu Huan and Polydoros, 1995; Hameed *et al.*, 2009] were developed providing optimal results, while other classifiers based on feature value were sub-optimal but computationally efficient. Some past AMC algorithms incorporated analog modulations but as the use of digital modulations dominated analog

modulations, recent works are more focused on digital modulations. AMC algorithms can be categorized based on the technique used or modulation type it considers. Classifiers in the thesis consider the former categorization. Two categories are broadly defined in literature: Likelihood Based and Feature-Based.

2.2.1 Likelihood-based Classifiers

Likelihood-based classifiers are multiple hypothesis testing problems in which the likelihood function for all modulation schemes considered are calculated and compared to give the final modulation classification decision. The likelihood function value tells about how likely the symbols belong to a well-defined ideal constellation of particular modulation. LB classifier gives the optimal decision if provided with the accurate channel information [Huan and Polydoros, 1995; Sills, 1999; Wei and Mendel, 2000; Hameed *et al.*, 2009; Ramezani-Kebrya *et al.*, 2013]. It provides optimal classification in the Bayesian sense, i.e., it minimizes the probability of error from the posterior distribution of possible constellations. LB methods struggle with high computational complexity as calculating the likelihood function is computationally complex and the Bayes classifier is difficult to implement. Different likelihood-based decisions are based on the ratio test between two hypotheses. Optimizing the thresholds for ratio test adds another level of complexity to these methods.

To combat unknown channel parameters and for the reduction in computation, likelihood functions are modified and maximum likelihood classifier, average likelihood ratio test, generalized likelihood ratio test, and hybrid likelihood ratio test were developed and discussed in this section.

Maximum likelihood classifier

The maximum likelihood classifier is one of the most popular classifiers. It calculates the posterior probability of every symbol corresponding to each modulation scheme with given PDF and channel parameters. All the parameters are known except for the modulation scheme. The likelihood value for each modulation scheme is compared and the decision is given in favor of the maximum value of likelihood. Depending on the channel model, the likelihood function takes different expressions.

Likelihood function in AWGN channel

The likelihood value for every signal point observed is equivalent to the probability of that point to be observed with the given channel model. For the AWGN channel, the probability that a point $d[n]$ is observed given the variance σ for modulation scheme $\mathbb{M}$ is given by

$$L(d[n]|\mathbb{M},\sigma) = p(d[n]|\mathbb{M},\sigma) \tag{2.12}$$

$$L(d[n]|\mathbb{M},\sigma) = \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|d[n]-A_m|^2}{2\sigma^2}} \tag{2.13}$$

Here A_m is a symbol point belonging to a particular modulation scheme from the modulation pool. The likelihood value is calculated corresponding to all modulation schemes with the channel parameters known. If total N symbols are observed, the joint probability for modulation scheme $\mathbb{M}$ is expressed as:

$$L(d[n]|\mathbb{M},\sigma) = \prod_{n=1}^{N} \sum_{m=1}^{M} \frac{1}{M} \frac{1}{2\pi\sigma^2} e^{-\frac{|d[n]-A_m|^2}{2\sigma^2}} \tag{2.14}$$

For calculation simplicity and comparison in maximum likelihood classifier, natural logarithm is used, i.e.,

$$logL(d[n]|M,\sigma) = log\left(\prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\sigma^2}e^{-\frac{|d[n]-A_m|^2}{2\sigma^2}}\right) \tag{2.15}$$

$$logL(d[n]|M,\sigma) = \sum_{n=1}^{N} log\left(\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\sigma^2}e^{-\frac{|d[n]-A_m|^2}{2\sigma^2}}\right) \tag{2.16}$$

Once the likelihood function is established according to the channel model and estimated channel parameters, its value is evaluated corresponding to each modulation scheme considered in the modulation pool. The decision of final modulation is given in favor of a modulation scheme corresponding to which, likelihood takes the highest value. M-PSK and M-QAM modulation can also be classified using ML methods considering their PDF phase [Shi and Karasawa, 2011, 2012].

Likelihood Ratio Test

To establish the likelihood function, information of exact channel parameters is a must. The likelihood function cannot handle any missing parameter. Either it is estimated, or the problem of a missing parameter is overcome using different likelihood ratio test [Panagiotou *et al.*, 2000]. In the average likelihood ratio test (ALRT) developed in 1990 [Polydoros and Kim, 1990], an unknown parameter in the likelihood function is integrated with all their possible values. This was further adopted in [Huan and Polydoros, 1995; Beidas and Weber, 1995; Hong and Ho, 2000]. For hypothesis H, ALRT likelihood function with unknown phase offset can be expressed as:

$$L_{ALRT}(r) = \int_{\theta_0} L(d|\theta_0)f(\theta_0|H)d\theta \tag{2.17}$$

$$= \int_{\theta_0}\prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\sigma^2}e^{-\frac{|d[n]-\alpha e^{-j\theta_0}A_m|^2}{2\sigma^2}}f(\theta_0|H)d\theta_0 \tag{2.18}$$

The final classification is done by using maximum likelihood value. ALRT involves the integration and it becomes more complex when the number of unknown parameters increases. Further, if an accurate model is not available, classification accuracy reduces. The generalized likelihood ratio test (GLRT) was developed to ameliorate the complexity [Panagiotou *et al.*, 2000]. In GLRT, instead of integrating, the likelihood function is maximized concerning unknown parameters over its possible range. For multiple unknown parameters, maximization is done over multiple parameters. Equation (2.20) shows the GLRT likelihood function considering phase offset as the unknown parameter.

$$L_{GLRT}(d) = max_{\Theta}L(d|\alpha,\sigma,\theta_0) \tag{2.19}$$

$$= max_{\Theta}\prod_{n=1}^{N}\sum_{m=1}^{M}\frac{1}{M}\frac{1}{2\pi\sigma^2}e^{-\frac{|d[n]-\alpha e^{-j\theta_0}A_m|^2}{2\sigma^2}} \tag{2.20}$$

Complexity in GLRT is reduced as compared to ALRT. However, for the modulation with the same structure and different orders like 4-QAM, 16-QAM; 16-QAM, 64-QAM; QPSK, 8-PSK, GLRT becomes a biased classifier. It is due to the overlapping of symbols with nested modulation schemes. Hence hybrid likelihood ratio test (HLRT) was developed. It is a combination of ALRT and GLRT where some parameters are handled by ALRT and others by GLRT. The complexity of likelihood is not increased while accuracy is also not much compromised. A discrete likelihood ratio test was developed to compute the likelihood function faster, which uses a lookup table quantized to a finite set for different modulation schemes [Xu *et al.*, 2010].

2.2.2 Feature Based Classification

Likelihood-based classifiers provide excellent accuracy but are computationally expensive. Feature-Based methods provide near-optimal performance with reduced computational complexity. Nowadays, digital modulations are more popular, but feature-based methods are used considering the analog modulation schemes as well. Feature-Based methods are divided into two steps; the first is the feature extraction, in which key features are extracted, and the second is pattern recognizer or classifier, which identifies modulation type based on the features extracted.

Feature-based classifiers extract the key features based on the signal's amplitude, phase, and frequency. These features differ according to modulation schemes and can be classified based on decision trees or any other classifier [Nandi and Azzouz, 1995; Azzouz and Nandi, 1995]. These days, classifiers based on machine learning tools are popular for their reliable performance. This section will cover all these features and classifiers for modulation classification.

Spectral-based feature

Spectral-based features exploit the unique characteristics of different signal modulations in three key signal aspects, namely the amplitude, phase, and frequency. The values of the extracted features are then used at different levels of decision trees for classification. The first feature, γ_{max}, is the maximum value of the spectral power density of the normalized and centered instantaneous amplitude of the recovered signal.

$$\gamma_{max} = max|DFT(A_{cn})|^2/N \tag{2.21}$$

Here DFT represents a Discrete Fourier Transform. The second feature, σ_{ap}, is the standard deviation (SD) of the absolute value of the non-linear component of the instantaneous phase.

$$\sigma_{ap} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n]>A_t} \phi_{NL}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n]>A_t} |\phi_{NL}[n]|\right)^2} \tag{2.22}$$

The third feature, σ_{dp} , is the SD of the non-linear component of the direct instantaneous phase.

$$\sigma_{dp} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n]>A_t} \phi_{NL}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n]>A_t} \phi_{NL}[n]\right)^2} \tag{2.23}$$

The fourth feature, P , is an evaluation of the spectrum symmetry around the carrier frequency.

$$P = \frac{P_L - P_U}{P_L + P_U} \tag{2.24}$$

$$P_L = \sum_{n=1}^{f_{cn}} |X_c[n]|^2$$

$$P_U = \sum_{n=1}^{f_{cn}} |X_c[n + f_{cn} + 1]|^2$$

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

$$\sigma_{af} = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n]>A_t} f_N^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n]>A_t} |f_N[n]|\right)^2} \tag{2.26}$$

σ_{aa}, σ_{af} , is SD of absolute value of normalized and centered instantaneous amplitude and frequency of the signal respectively.

$$f_N[n] = f_m[n]/f_s$$

$$f_m[n] = f[n] - \mu_f$$

$$\mu = \frac{1}{N}\sum_{n=1}^{N} f[n]$$

σ_a is SD of normalized and centered instantaneous amplitude.

$$\sigma_a = \sqrt{\frac{1}{N_c}\left(\sum_{A_n[n]>A_t} a_{cn}^2[n]\right) - \left(\frac{1}{N_c}\sum_{A_n[n]>A_t} a_{cn}[n]\right)^2} \tag{2.27}$$

$$\mu_{42}^a = \frac{E\{A_{cn}^4[n]\}}{\{E\{A_{cn}^2[n]\}\}^2} \tag{2.28}$$

$$\mu_{42}^f = \frac{E\{f_N^4[n]\}}{\{E\{f_N^2[n]\}\}^2} \tag{2.29}$$

μ_{42}^a, μ_{42}^f are kurtosis of the normalized and centered instantaneous amplitude and frequency. These features are compared with an optimized threshold to classify between FM, M-FSK, DSB, SSB, M-PSK, M-QAM, AM, M-ASK, and VSB.

Wavelet Based features

Wavelet transform of a function $y(t)$ with wavelet function $\psi_{a,\tau}$ is expressed as:

$$CWT(a,\tau) = \int_{-\infty}^{\infty} y(t)\psi_{a,\tau}^*(t)dt \tag{2.30}$$

Let the expression for ASK, PSK, FSK, and QAM modulation schemes are given by:

$$y_{ASK}(t) = \sum_{n=1}^{N} |A_n| p_{t_s}(t - nt_s) \tag{2.31}$$

$$y_{FSK}(t) = \sqrt{S}\sum_{n=1}^{N} e^{j(\omega_n t + \phi_0)} p_{t_s}(t - nT_s) \tag{2.32}$$

$$y_{PSK}(t) = \sqrt{S}\sum_{n=1}^{N} e^{j\phi_n} p_{T_s}(t - nT_s) \tag{2.33}$$

$$y_{QAM}(t) = \sum_{n=1}^{N} A_n p_{T_s}(t - nT_s) \tag{2.34}$$

In the above expressions, A_n is symbol amplitude, ω_n is signal frequency for n_{th} symbol, T_s is symbol time, ϕ_n is phase of n_{th} symbol, and p_{T_s} is pulse shaping filter. If the wavelet function used is expressed as:

$$\psi(t) = \begin{cases} 1, & \text{if } 0 \le t < \frac{T}{2}; \\ -1, & \text{if } \frac{T}{2} \le t < T; \\ 0, & \text{otherwise.} \end{cases}$$

Wavelet transform of different modulation schemes using Haar wavelet is expressed as:

$$|CWT_{ASK}(a,\tau)| = \frac{4|A_n|}{\omega_c\sqrt{a}} \sin^2\left[\omega_c \frac{aT_s}{4}\right] \tag{2.35}$$

$$|CWT_{FSK}(a,\tau)| = \frac{4\sqrt{S}}{(\omega_c+\omega_n)\sqrt{a}}\sin^2\left[\frac{(\omega_c+\omega_n)aT_s}{4}\right] \tag{2.36}$$

$$|CWT_{PSK}(a,\tau) = \frac{4\sqrt{S}}{(\omega_c\sqrt{a}}\sin^2\left[\omega_c\frac{aT_s}{4}\right] \tag{2.37}$$

$$|CWT_{QAM}(a,\tau) = \frac{4|A_N|}{\omega_c\sqrt{a}}\sin^2\left[\frac{aT_s}{4}\right] \tag{2.38}$$

PSK has a constant value of CWT, while others have multi-step functions. After normalizing the CWT, FSK is differentiated based on the variance in CWT. ASK and QAM are not distinguished using CWT; hence higher-order statistics were used.

High-order Statistics based features

Initially, the third-order moment was used on a demodulated signal [Hipp, 1986]. It was extended to higher-order moments applied to the signal phase to classify M-PSK modulation. Higher-order moments for M-PSK classification were used in [Soliman and Hsue, 1992]. Higher-order cyclic moments were used in [Spooner, 1996] to classify modulation with similar cyclic autocorrelation functions. m^{th} order moment of signal phase is given by:

$$\mu_k(d) = \frac{1}{N}\sum_{n=1}^{N}\phi^k(n) \tag{2.39}$$

Here ϕ_n is the phase of n^{th} signal sample. For the classification of M-PAM, M-PSK, and M-QAM modulations, different order cumulant-based features were used. Second and fourth-order cumulants [Swami and Sadler, 2000] are expressed as:

$$\hat{C}_{20} = \frac{1}{N}\sum_{n=1}^{N} r^2[n] \tag{2.40}$$

$$\hat{C}_{21} = \frac{1}{N}\sum_{n=1}^{N} |r[n]|^2 \tag{2.41}$$

$$\hat{C}_{40} = \frac{1}{N}\sum_{n=1}^{N} r^4[n] - 3\hat{C}_{20} \tag{2.42}$$

$$\hat{C}_{41} = \frac{1}{N}\sum_{n=1}^{N} r^3[n]r^*[n] - 3\hat{C}_{20}\hat{C}_{21} \tag{2.43}$$

$$\hat{C}_{42} = \frac{1}{N}\sum_{n=1}^{N} |r[n]|^4 - |\hat{C}_{20}|^2 - 2\hat{C}_{21}^2 \tag{2.44}$$

These values differ for different orders of M-PAM, M-PSK, and M-QAM. A proper decision tree with an optimized threshold can differentiate these modulation schemes.

Other Features

Apart from spectral-based features, wavelet-based features, different modulations show differences in cyclic spectrum patterns, which is used as a feature for modulation classification. This is called cyclostationarity property [Gardner and Spooner, 1988; Ramkumar, 2009].

Other modulation classification method includes the signal cumulative distribution assuming channel parameters are known. The ideal signal distribution of symbol mapping is known as apriori. Modulation is classified as the one which matches the underlying

distribution the most. This similarity matching process is called the goodness of fit test. The Kolmogorov-Smirnov (K-S) test was first used by wang and wang considering this method less complex than the likelihood-based classifier [Wang and Wang, 2010; Wang and Chan, 2012; Wang *et al.*, 2016]. Carmer-Von Mises Test, Anderson- Darling test is also used as the goodness of fit test.

2.3 MACHINE LEARNING BASED CLASSIFIERS

In the previous section, different features were explained. These features are given to a decision tree where they are classified after comparison with the threshold. At each stage, these thresholds need to be optimized. Further, each feature is considered at a time in the hierarchy which adds another level of optimization. Hence different machine learning tools are used which optimizes the thresholds, considers all features at a time, and is easy to implement. Different features do not play a role in classification as their values overlap for different modulation schemes. Different machine learning methods like genetic programming, linear regression, artificial neural networks are used to reduce the dimension of the feature set, which lowers the computational burden [Nandi and Azzouz, 1998].

ML classifiers can be based on supervised learning or unsupervised learning. In supervised learning-based classifiers, modulated signals or features extracted are used to train the classifier for weight or threshold optimization. Different supervised classifiers include Artificial Neural networks (ANN), k-NN classifiers [Aslam *et al.*, 2012, 2010], logistic regression, naive Bayes classifiers [Wong *et al.*, 2008], support vector machine [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], etc.

ANN was first used by Azzouz and Nandi for modulation classification [Azzouz and Nandi, 1995; Nandi and Azzouz, 1998]. In [Nandi and Azzouz, 1998], spectral-based features were used for training and testing of ANN to classify different analog and digital modulations. ANN was also used with like cyclic spectral-based features [Lu Mingquan *et al.*, 1996], instantaneous phase, and frequency-based features [Hossen *et al.*, 2007], or a combination of features [Wong and Nandi, 2004]. With increasing computational power in hand, the number of hidden layers was increased, which leads to further improved classification accuracy due to capturing of minute features from the input and leads to the new area of research popularly called deep neural networks (DNN) [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; Wang *et al.*, 2020a,b; Ma *et al.*, 2020].

DNN extracts and learns the significant features without any feature engineering. Different input features like cumulants and moments [O'Shea *et al.*, 2018], a color image formed using constellation [Peng *et al.*, 2018], complex baseband symbols with centroids estimated using fuzzy c-means [Ali *et al.*, 2017] are used to train DNN. Different developed models like ResNet-50, Inception ResNet V2 are used in a hierarchical fashion for efficient modulation classification [Kumar *et al.*, 2020].

In unsupervised learning-based classifiers, constellation shape, graph-based methods, and clustering algorithms on constellations are used for modulation classification. Different clustering algorithms like subtractive clustering, hierarchical clustering [Swami and Sadler, 2000], fuzzy c means [Ahmadi, 2010], k-medoids clustering are used on feature sets or constellations for modulation classification [Jajoo *et al.*, 2018].

…