PERFORMANCE ANALYSIS OF HIERARCHICAL CLASSIFICATION OF MODULATION TYPES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY

BY

BENGİSU YALÇINKAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

JANUARY 2020

Approval of the Graduate School of Natural and Applied Sciences, Atilim University.

Prof. Dr. Ali Kara

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of **Master of Science in Electrical and Electronics Engineering Department, Atilim University**.

Assoc. Prof. Dr. Kemal Efe Eseller

Head of Department

This is to certify that we have read the thesis PERFORMANCE ANALYSIS OF HIERARCHICAL CLASSIFICATION OF MODULATION TYPES submitted by BENGİSU YALÇINKAYA and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Prof. Dr. Ali Kara

Supervisor

Examining Committee Members:

Assoc. Prof. Dr. Enver Çavuş

Asst. Prof. Dr. Beytullah Yıldız

Asst. Prof. Dr. Özgür Ergül

Asst. Prof. Dr. Hakan Tora

Prof. Dr. Ali Kara

Date: January 22, 2020

I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : BENGİSU YALÇINKAYA

Signature :

ABSTRACT

PERFORMANCE ANALYSIS OF HIERARCHICAL CLASSIFICATION OF MODULATION TYPES

Yalçınkaya, Bengisu
M.S., Department of Electrical and Electronics Engineering
Supervisor : Prof. Dr. Ali Kara

January 2020, 58 pages

Automatic modulation classification (AMC) is a frequently required framework to determine the modulation type of an incoming modulated signal with an unknown modulation type. AMC applications are divided under two main titles in the literature as likelihood-based (LB) and feature-based (FB) methods. In this thesis, an AMC algorithm is developed with a FB approach. As classifier, Support Vector Machine (SVM) using linear, quadratic and cubic kernel is chosen and their performances are compared. Over-the-air collected modulated signals with the SNR values between 0 and 30 dB are used. Signals are modulated with 12 different digital modulation types containing M-ASK, M-PSK, M-APSK up to higher orders. Statistical features i.e. mean, variance, skewness and kurtosis of the instantaneous amplitude, phase and frequency of the signal are used in addition to higher-order moments and cumulants up to 8^{th} order. SVM using quadratic kernel showed slightly higher performance. In addition, a hierarchical classification structure with less complexity compared to the literature has been proposed in order to improve performance especially in high order modulation types which show very poor performance when classified with using a single classifier. A significant improvement is observed in the accuracies of these modulations comparing with the traditional method. The overall performance is increased from 80% to 90%.

Keywords: Automatic modulation classification, feature extraction, digital modulation, higher order statistics, support vector machines

ÖZ

MODÜLASYON TÜRLERİNİN HİYERARŞİK SINIFLANDIRILMASININ PERFORMANS ANALİZİ

Yalçınkaya, Bengisu
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği
Tez Yöneticisi : Prof. Dr. Ali KARA

Ocak 2020, 58 sayfa

Otomatik modülasyon sınıflandırması (AMC), bilinmeyen bir modülasyon tipine sahip gelen modüle edilmiş bir sinyalin modülasyon tipini belirlemek için sıklıkla ihtiyaç duyulan bir yapıdır. AMC uygulamaları literatürde olabilirlik tabanlı (LB) ve özellik tabanlı (FB) yöntemler olarak iki ana başlık altında bölünmüştür. Bu tezde, FB yaklaşımı ile bir AMC algoritması geliştirilmiştir. Sınıflandırıcı olarak lineer, kuadratik ve kübik çekirdek kullanan Destek Vektör Makinesi (SVM) seçilmiş ve performansları karşılaştırılmıştır. SNR değerleri 0 ila 30 dB arasında olan havadan toplanan modüle edilmiş sinyaller kullanılmıştır. Sinyaller, yüksek derecelere kadar M-ASK, M-PSK, M-APSK içeren 12 farklı dijital modülasyon tipiyle modüle edilmiştir. İstatistiksel özellikler, yani sinyalin anlık genliği, fazı ve frekansının ortalaması, varyansı, çarpıklığı ve basıklığı, 8. dereceye kadar olan daha yüksek dereceli momentlere ve kümülanlara ek olarak kullanılmıştır. Sınıflandırıcılar arasından ikinci dereceden çekirdek kullanan SVM daha yüksek performans göstermiştir. Ayrıca, özellikle tek bir sınıflandırıcı kullanılarak sınıflandırıldığında çok düşük performans gösteren yüksek dereceli modülasyon tiplerinde, performansı arttırmak için literatüre kıyasla daha az karmaşıklığa sahip bir hiyerarşik sınıflandırma yapısı önerilmiştir. Bu modülasyonların doğruluklarında geleneksel yönteme kıyasla önemli bir gelişme gözlenmektedir. Genel performans

%80'den %90'a yükselmiştir.

Anahtar Kelimeler: Otomatik modülasyon sınıflandırma, özellik çıkarımı, sayısal modülasyon, yüksek mertebeden istatistikler, destek vektör makinaları

To the ones who never give up and never let me give up...

ACKNOWLEDGMENTS

First of all, I would like to express my deep and sincere gratitude to my supervisor Prof. Dr. Ali Kara for providing his valuable guidance and tolerance. His endless patience and extensive knowledge led me to realize this work. Working with him was a great privilege and luck since he taught me how to learn and research above all.

I shall also thank to Asst. Prof. Dr. Hakan Tora specially, for his valuable comments and suggestions. He guided me at a deadlock moment and allowed me to look through a new window.

Furthermore, I thank the rest of the members of the thesis committee: Assoc. Prof. Dr. Enver Çavuş, Asst. Prof. Dr. Beytullah Yıldız and Asst. Prof. Dr. Özgür Ergül for enlightening my future work by bringing a wider perspective to my work with precious comments and questions.

My sincere thanks to Remziye Büşra Tezel, my dearest colleague and heartfelt friend for being a companion in this journey, all the fun we have had and giving me hope when I was desperate.

A huge thanks to Gökhan Gökdoğan one of the greatest builder of my academic career, has always believed in me and supported me all over for years. He has been an emotional punching bag and cry pillow yet never let me give up.

Last but not least, I would like to express my endless thanks to my family, who has always been there for me in all my successes and failures.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF SYMBOLS

α_n	:	Instantaneous Amplitude of a Complex Signal
ϕ_n	:	Instantaneous Phase of a Complex Signal
f_n	:	Instantaneous Frequency of a Complex Signal
μ	:	Mean of a Complex Signal Characteristic
σ^2	:	Variance of Complex Signal Characteristic
γ	:	Skewness of Complex Signal Characteristic
κ	:	Kurtosis of Complex Signal Characteristic
M_{pq}	:	Moment of a Complex Signal
C_{pq}	:	Cumulant of a Complex Signal

CHAPTER 1

INTRODUCTION

Communication, in general, is a transmission of information over any medium. Information generally consists of image, voice or any type of data. Medium is named as channel and can be formed of either cable or air. A communication system as a whole mainly consists of a transmitter, a receiver and a channel through which information is transmitted from the transmitter to the receiver. The message signal produced by a source of information is processed by the transmitter and made ready for transmission over a channel. As the signal passes through the channel, disturbances in the signal occur due to noise, channel imperfections and interferences from other signals. In the end, the distorted signal received by the receiver is processed again to obtain the original message [1]. Received signal can simply be expressed as $r(t) = s(t) + n(t)$ where $s(t)$ is considered as the transmitted signal, $n(t)$ is the noise and $r(t)$ is the received signal at a given time t.

As mentioned above, the information from the transmitter to the receiver can be transmitted by electromagnetic waves over a wireless medium as well as by using a cable. Communication systems that use a wireless medium for transmission are called wireless communication. There are a variety of channel effects that affect the transmitted signal. Additive white gaussian noise (AWGN) is considered to be the most fundamental and simplest modelable channel effect. In the AWGN channel model, the signal is distorted by adding a white noise with Gaussian distribution. This channel model is insufficient to model real-life conditions, as other disturbances encountered in real life are not taken into account. In addition to AWGN, multipath fading channel types which have much more complexity such as Rayleigh and Rician fading are used to approach real-life conditions. If there isn't a line of sight (LOS) between the transmitter and the receiver due to buildings, walls or other obstacles then the channel can be modelled as Rayleigh fading channel type. If there is a LOS between the transmitter and the receiver in addition to reflections from surroundings then the Rician fading channel type can be used for modelling. The more distortions in the signal, the more difficult it is to regain that signal [2]. In communication systems, the ratio of signal power to noise power is represented by the signal-to-noise ratio (SNR) concept as $SNR = P_{signal}/P_{noise}$.

Depending on the type of the message to be sent, communication systems are grouped under two main headings, namely analog communication and digital communication. Signals that are continuously defined over a time interval and can take an infinite number of values in that interval are defined as analog signals. Signals that are defined only during certain periods of a finite time interval and can only take a certain set of values in that interval are defined as digital signals. Compared to analog communication, digital communication is the most widely used communication technique today due to the fact that it provides more amount of data transmission with less error, has high noise immunity and is easy to process. Unlike analog communication, digital communication does not need to fully recover the original signal from the distorted signal at the receiver, it is sufficient to determine which of the possible waveforms is. For instance, in binary signals, there are two possible values that the waveform can take like 0 and 1 but an analog signal can take infinite possible values. Which is why digital signals are more tolerated to interferences and distortions [3].

In both analog and digital communication, the message signal should be modified to propagate through the channel. This modification is carried out by means of the modulation concept. The bandpass transmission of the signals is provided with the help of antennas. Generally, the message signals to be sent are of high wavelength and low frequency and therefore low energy signals. The relationship between the wavelength and the frequency can be described as $\lambda = v/f$ where v is the velocity of the wave in a medium (in free space, it is the speed of light c), f is the frequency and λ is the wavelength. The wavelength and the antenna length are directly proportional to each other [4]. Although the fading effect at lower frequencies is less, in order to transmit such high wavelength signals over long distances, antenna dimensions that cannot be realized are required. Within this constraint, the transmitted signals must be superimposed on a higher frequency signal to complete their inter-antenna travel. The digital or low-frequency signal containing the information to be transmitted, called the modulating signal, modulates some of the properties such as amplitude, phase or frequency of the high-frequency signal called the carrier signal. This process is called modulation and the device that performs this process is called modulator. When the signal is received by the receiver, the original signal is extracted from the carrier signal. This process is called demodulation and the device that performs this process is called demodulator. In Figure 1.1, a simplified communication system from the source to the sink is illustrated. The devices called modems, which we can encounter everywhere in our daily lives, can perform both modulation and demodulation. There are many other advantages of transmitting signals at a higher frequency rather than a low frequency, such as immunity of interference and noise also extended transmission range.

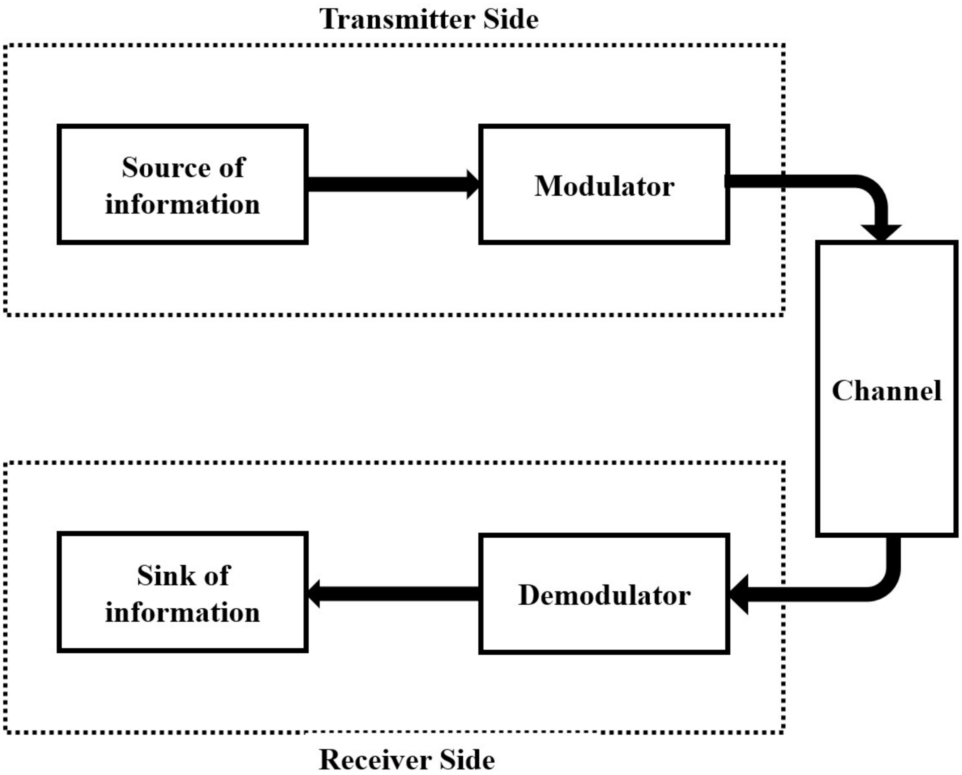

Figure 1.1: A Simplified Communication System

Modulation techniques are basically divided into two groups as analog modulation and digital modulation. Analog modulation is the general name given to the modulation techniques used to transmit analog signals described above, such as the audio signal. The most commonly used analog modulation techniques are amplitude modulation (AM), frequency modulation (FM) and phase modulation (PM). In AM, the amplitude of the analog carrier signal is modified according to the amplitude of the message signal. In PM and FM which are called angle modulation, phase and frequency vary with the same logic. Digital modulation is used to transmit the digital signals described above such as digital bitstreams. Amplitude-shift keying (ASK), phase-shift keying (PSK), frequency-shift keying (FSK), Amplitude-Phase shift keying (APSK) and quadrature-amplitude modulation (QAM) are the most commonly used types of digital modulation. In digital modulation, as in analog modulation, ASK is the amplitude variation, FSK is the frequency variation, and PSK is the phase variation of the analog carrier. The difference is that the modulating information signal is digital. A simplified table for some fundamental modulation types is given in Figure 1.2.

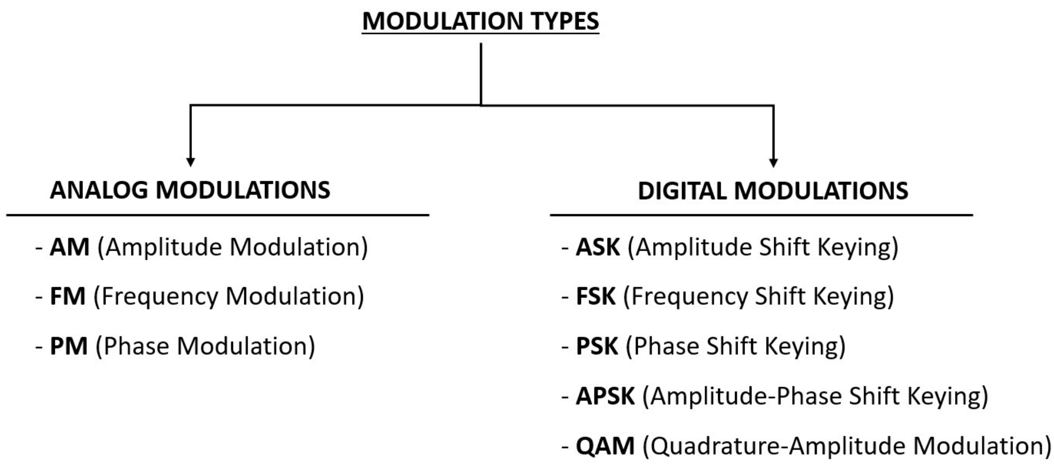

Figure 1.2: Fundamental Modulation Types

In communication systems, the main purpose is to obtain the signal transmitted by the transmitter with minimum loss and error rate on the receiving side. The most important parameter to achieve this objective is to demodulate the signal with a demodulator suitable for the modulation type in which the signal is modulated. The receiver side must, therefore, know the modulation type of a signal modulated by the transmitter [5]. In such cases where it is not possible to know the modulation type of the received modulated signal, modulation classification (MC) methods are used widely. Signal processing applications in communication systems are mostly built on a foundation of statistical basis. The MC problem can also be summarized simply by a $P(X/Y)$ conditional probability function. Assuming that Y is a received signal, X can be called any modulation scheme. Initially, the MC process was performed by engineers manually by examining the signals and their characteristics one by one. Later on, modern communication techniques were developed in order to perform these processes automatically and these were generally called automatic modulation classification (AMC). In a communication system, the AMC has found a place between the signal detection and demodulation steps. With AMC, the modulation type of the received modulated signal is automatically found then the appropriate demodulator is selected and the original data sent is obtained. Nowadays, it is of great importance, especially in the military and commercial industries, to classify such signals which have unknown modulation type quickly and automatically with minimal error [6]. It is possible to see military applications in surveillance and electronic warfare and commercially cognitive radio applications [7], [8]. Channel quality has an enormous effect on AMC. This problem have been tried to be solved from many different perspectives in the literature such as orthogonal frequency division multiplexing (OFDM) applications [9], especially in channel types such as multipath fading where the transmitted signal can be disturbed highly [10]. Today, the ability to easily handle different modulation types using software-defined radios (SDR) as hardware has also played a major role in the development of AMC [11]. In addition, classification of the signals from multiple antennas i.e. multiple-input multiple-output (MIMO) systems are found a place in the literature [12]. Recently, in adaptive transceiver applications where the transmitter automatically changes the modulation type according to the channel type, AMC eliminates the need to send modulation information along with the signal and spend bandwidth redundantly [13].

The optimal decision criteria and signal characteristics for AMC have been studied for many years. AMC techniques are categorized under two main titles traditionally as likelihood-based (LB) and feature-based (FB) approaches [14].

The LB approach selects the modulation type which has the maximum likelihood from the available modulation schemes with a multi-hypothetic process [15]. With the assumption of knowing the probability density function (pdf) of the received signal, a maximum likelihood ratio test is used with a likelihood function to make a decision by comparing with a threshold value [16], [17]. A variety of LB applications have been developed, such as the average likelihood ratio test (ALRT) [18], the generalized likelihood ratio test (GLRT) [19], and the hybrid likelihood ratio test (HLRT) which is the combination of previous two methods [20]. Usually, LB algorithms gives optimal results according to other approaches. However, it has been seen in the literature that the performance of LB algorithms is not high in some special cases which have model mismatches with frequency offset and requiring noise immunity. In addition, the computational complexities of LB algorithms are quite heavy [21].

In the FB approach, various features are extracted that are specific to each different signal type, which defines and characterize each of them. These features are employed to distinguish between different signal types [22]. There are many features based on the different principles used for AMC applications. Basically, instantaneous amplitude, phase and frequency [23], Fourier and wavelet transform [24], spectrum symmetry [25], signal constellations [26], higher-order moments (HOMs) and cumulants (HOCs) [27], higher-order cyclic cumulants [28], very higher-order statistics (VHOS) have been widely used as features in the literature. Statistical features i.e HOMs and HOCs are generally referred to as higher-order statistics (HOS). Instantaneous amplitude, phase and frequency are the statistical characteristics of the signal, they don't usually show high performance at low SNR values. Fourier and Wavelet transform features can be extracted by transforming the signal to these domains. Wavelet transform uses transient differences of modulated signals, they cannot show high success in classifying modulated signals according to their modulation orders. Cyclostationary features are usually used for linear classification of modulation types at low SNR levels [29]. When HOS are used as features, they can perform well in low SNRs because of their high noise immunity. Also when HOCs are used, the classification of higher-order modulation types becomes more possible. The most important parameter in the FB approach is the proper selection of features. The type of modulation plays the most important role in determining the appropriate features [30]. Although FB approach does not give the optimal solution, LB and FB approaches do not show noticeable differences in success performance when appropriate features are used. Moreover, FB applications are more advantageous in terms of computational complexity, robustness and easy to implementation than LB applications.

The extracted features are input to a suitable classifier to distinguish the signals of different modulation types from each other. There are many classifiers with different process times and complexities with varying performances for several situations such as different channel effects. Classification is generally done according to a threshold or using a pattern recognition method [31]. There are also studies where the decision-theoretic approach is applied in cases where the carrier signal is fully known [32] however AMC becomes a challenging issue when there is no a priori information about the incoming signal. Support vector machines (SVM) [33], decision trees [34], k-nearest neighbours (KNN) [35], hidden Markov model (HMM) and artificial neural networks [36] are some examples of them. Hierarchical structures can be used to distinguish modulations according to their orders with various methods such as using thresholds [27], polynomial classification [37] or neural networks [38]. It has been observed in the literature that this method consists of many classifiers and stages. In general, the higher the complexity of the classifier, the higher the success rate. The main challenge here is to achieve low complexity and high success trade-offs.

Machine learning (ML) and deep learning (DL) algorithms are frequently used as classifiers in the literature. The use of machine learning (ML) algorithms to solve various problems encountered in wireless communication networks is becoming increasingly common. ML applications in communication systems are used especially in quantization, channel modelling, equalization [39] and MC problems for both digital and analog modulations [40]. ML implementations mainly are evaluated in two different categories as supervised learning and unsupervised learning in the literature. When using supervised learning algorithms in classification problems, the dataset is usually divided into two. The first part of the dataset is used to train the algorithm, while the other part is used to test the trained algorithm. A class label is given to each data in the dataset used for training. In the testing process, the classifier is expected to seperate test set according to these class labels. In unsupervised learning such as clustering

[41], when the dataset is given to the algorithm, the class labels are not given with the data and the algorithm is expected to learn the characteristics and hidden patterns of the data and group the data according to these characteristics [42].

SVMs, decision trees, KNNs and neural networks (NN) are frequently used as classifiers in supervised ML applications. SVM which have non-parametric model, was showed up for binary classification, later has been developed and applied in multi-class classification and regression processes. After the feature extraction process, the SVM method is applied to the created feature space and classification process can be performed successfully. SVM, which takes shape according to the kernel function, has many different derivations such as linear, polynomial and gaussian radial.

Decision trees classify by using the proximity of the data in the sample space. The decision tree is formed by the separation of different conditions like tree branches. Data continues its path in the decision tree by selecting the branch closest to its properties. On the other hand, the KNN classifier performs the classification process taking into account the distance of the samples in the feature space to the neighboring samples.

Another learning technique that has recently become widespread in communication systems is DL [43]. DL methods have found wide coverage in the literature in signal processing, especially image [44] and voice [45] processing, and computer vision applications. Recently, there are examples of DL methods for estimating modulation types and attempting to classify data. Comparisons were made with other methods and classifiers, such as FB, by giving the raw data itself to the network instead of the features extracted from the data [46]. This learning method is developed by using algorithms consisting of multi-layer artificial neural networks. NN algorithms allow data that cannot be separated by linear classifiers to be classified as non-linear with the help of an activation function. Normally, the matrix product is used to calculate this activation function, while convolutional NN uses the convolution function to calculate the activation function [47]. In the deep neural network, the output of each layer can be assumed to be the input of the next layer, thus increasing the number of layers and therefore the complexity. In each layer, the inputs are weighted with certain weight matrices. In the Residual Network, the output of one layer can be the input of the next two layers. This makes it possible to go deeper in the network. Although it has not yet gained a large place in MC applications, it has led to improvements in image and voice processing applications.

In this thesis, over-the-air collected radio signals which modulated with various modulation types are classified with a FB approach by using various classifiers and their performances are compared. The classification process is based on machine learning applications. As a classifier, SVM and its polynomial kernel derivations up to 3^{rd} degree have been selected which outperforms compared to other ML-based classifiers. Mean, variance, skewness and kurtosis values of the instantaneous amplitude, phase and the frequency are used as statistical features. Additionally, HOMs and HOCs up to 8^{th} order are used to classify especially high order modulation types. Firstly, the performance of linear, cubic and quadratic SVM classifiers at different SNR values is examined using these features. Then, a hierarchical classification structure is proposed to increase the performance of the classification process, especially in high order modulations.

After an overview of the MC domain along with the introduction part, the rest of the thesis is organized as follows; In the second chapter, the introduction of the signal model and characteristics are given. Then the modulation types used in this thesis are presented. After that, the statistical features, their extraction methods and the classification algorithms are provided. Chapter 3 is devoted to the results and the performance analysis of the classifiers. In the last chapter 4, the results are discussed, flaws or shortcomings are determined and the conclusions are drawn.

CHAPTER 2

METHODOLOGY

There are two key steps to be considered in feature-based (FB) automatic modulation classification (AMC) algorithms. The first step is the proper selection and extraction of features according to the modulation types with using signal processing algorithms and estimations of signal properties such as amplitude, phase, frequency and noise power. The second one is applying the classifier in accordance with the features, modulation schemes, channel conditions and signal types. The selection of suitable features is critical for the classifier to work properly.

In this chapter, the methods used to construct an AMC structure in this thesis are addressed in all aspects. First of all, the signal model is introduced and the information about the characteristics of signals is given and the preprocessing implementations of the signals are mentioned. Then in the modulation types subsection, information about the modulation types in the data set is given. In the feature extraction subsection, the features used in this study are defined and the methods of obtaining them are discussed. Finally, classification subsection is dedicated to the supervised learning-based classifiers used in this study and their applications.

2.1 Signal Model and Characteristics

In real-world, modulation and demodulation processes usually operates in the form of complex signal notations. The complex form of the sampled modulated signal in the time domain which is the output of a Hilbert filter is given in Equation 2.1. With

Hilbert transform, analytic representation of the signal is achieved.

$$x(n) = I(n) + jQ(n) \tag{2.1}$$

where $I(n)$ is the instantaneous in-phase which is real and $Q(n)$ is the instantaneous quadrature-phase which is the imaginary component of the complex signal $x(n)$. The real part and the imaginary part are each other's Hilbert transforms.

The instantaneous signal characteristics i.e. instantaneous amplitude $a(n)$, phase $\phi(n)$ and frequency $f(n)$ which are useful for feature extraction are given in Equation 2.2, 2.3 and 2.4 respectively;

$$a(n) = |x(n)| = \sqrt{I^2(n) + Q^2(n)} \tag{2.2}$$

$$\phi(n) = tan^{-1}[\frac{Q(n)}{I(n)}] \tag{2.3}$$

$$f(n) = \frac{1}{2\pi} \frac{\phi(n) - \phi(n-1)}{\Delta n} \tag{2.4}$$

The mean values of these instantaneous signal characteristics are subtracted from them in order to eliminate any bias that may arise from the data collection system and manipulate results. In this manner, each of these characteristics is ”centered”. The centering process is conducted for $a(n)$ and $\phi(n)$ as shown in Equation 2.5 and 2.6 respectively assuming N is the total number of sample and n=1,2,...,N.

$$a_c(n) = a(n) - \mu_a \tag{2.5}$$

$$f_c(n) = f(n) - \mu_f \tag{2.6}$$

where μ_a is the mean value of $a(n)$ and μ_f is the mean value of $f(n)$ computed from N samples. Apart from the amplitude and the frequency, when performing the centering process for $\phi(n)$, firstly the linear component is subtracted because it may consist of an incorrect estimate of the frequency at the down-conversion stage [48]. Then the mean value of the non-linear phase $\mu_{\phi_{nl}}$ is removed as in the amplitude and frequency. Equation 2.7 and 2.8 shows the subtraction of the linear component and the mean of $\phi(n)$ respectively;

$$\phi_{nl}(n) = \phi(n) - 2\pi\mu_f(n)\Delta_t \tag{2.7}$$

$$\phi_{cnl}(n) = \phi_{nl}(n) - \mu_{\phi_{nl}} \tag{2.8}$$

where Δ_t is the duration between time samples, $\phi_{nl}(n)$ is the non-linear component of the phase and $\phi_{cnl}(n)$ is the centered non-linear phase.

2.2 Modulation Types

In this thesis, signals modulated with OOK, 4-ASK, 8-ASK, BPSK, QPSK, 8-PSK, 16-PSK, 32-PSK, 16-APSK, 32-APSK, 64-APSK and 128-APSK modulation types are used to extract features and construct a classification structure. All of these modulation types are digital up to higher-order as well as depend on amplitude and/or phase variations. The modulation types and their orders taken in the consideration in this study are listed in Figure 2.1.

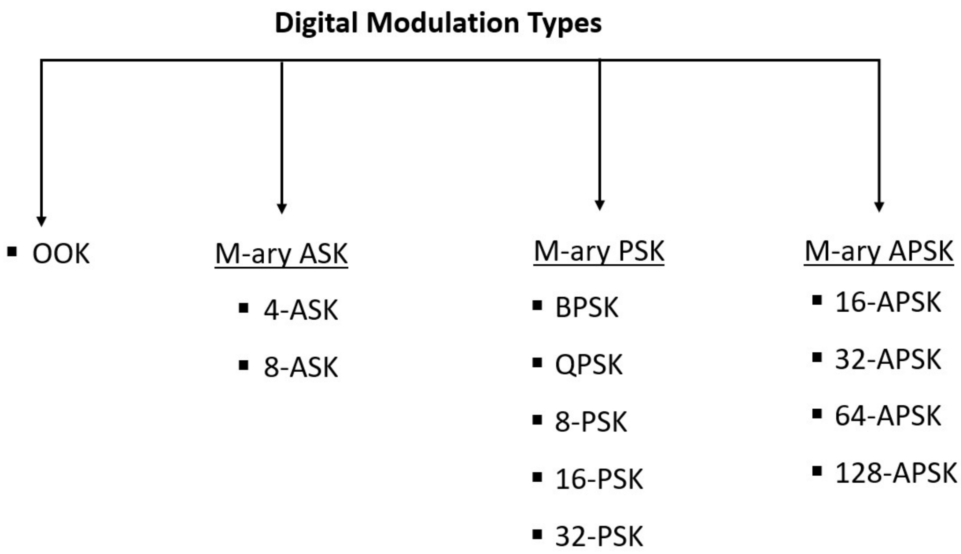

Figure 2.1: Digital Modulation Types

In the modulation types described in the following subsections, the amplitude value is shown in terms of energy in the analytical representation of the modulated signals since it is more convenient to calculate the error rate. The derivation of this expression is given in Equation 2.9 to 2.12;

$$s(t) = Acos(wt) \tag{2.9}$$

where A is the amplitude and it can be written in terms of root-mean-square value,

$$s(t) = \sqrt{2A_{rms}^2}cos(wt) \tag{2.10}$$

here A_{rms}^2 is equal to average power P,

$$s(t) = \sqrt{2P}cos(wt) \tag{2.11}$$

Power in *watts* can be represented as energy (E) in $joule/T$,

$$s(t) = \sqrt{\frac{2E}{T}}cos(wt + \phi) \tag{2.12}$$

2.2.1 On-Off Keying (OOK)

One of the digital modulation method On-Off Keying (OOK) is obtained by modulating the amplitude of a carrier waveform by a digital information signal which contains bits. In OOK which is also a version of amplitude-shift keying (ASK), the presence of the carrier signal is represented when the message signal is binary 1, and the absence of the carrier signal when the message signal is binary 0. The illustration of the OOK modulated signal in the time domain is given in Figure 2.2 [49].

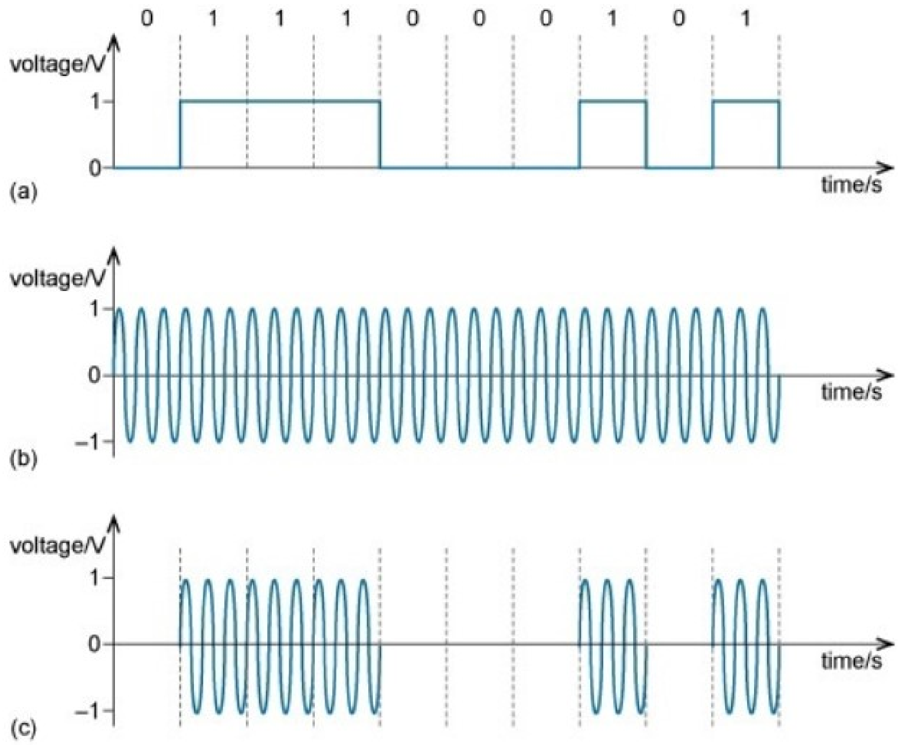

Figure 2.2: a) Message Signal, b) Carrier Signal, c) Modulated Signal for OOK

The general equation for OOK modulated signal $s(t)$ is given in Equation 2.13;

$$s_i(t) = \sqrt{\frac{2E_i(t)}{T}} cos(w_0 t + \phi) \tag{2.13}$$

where $0 \leq t \leq T$, $i = 1, 2$, $E_i(t)$ is the energy of the symbol at i^{th} level, $\sqrt{\frac{2E_i(t)}{T}}$ represents the amplitude and can take M different values, T is the symbol duration, w_0 is the angular frequency and ϕ is a constant phase. Amplitude is equal to zero if the message signal to be transmitted is binary 0.

The constellation diagram represents an indication of the amplitude and phase values of the signal and shows all available symbol values. The distance from the origin gives the amplitude value and counterclockwise from the x-axis gives the phase shift. The constellation diagram of the OOK signal is given in Figure 2.3

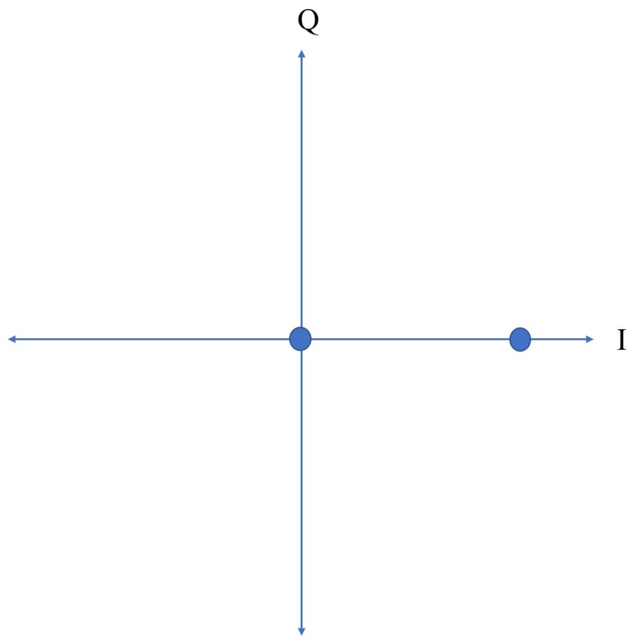

Figure 2.3: Constellation Diagram of OOK

where the x-axis represents in-phase and the y-axis represents quadrature value of the signal. As can be seen from the figure, there are two possible waveforms for OOK.

2.2.2 M-ary Modulations

In digital modulation techniques, if multiple bits are transmitted at one time to effectively use bandwidth, this is called M-ary modulation. M is the number of possible waveforms available for the modulation. A collection of bits that constitutes a piece of message information is called a symbol. The number of bits to be transmitted N for a given M can be found as shown in Equation 2.14.

$$N = \log_2 M \tag{2.14}$$

The symbols in the message signal can be considered as each element of an alphabet with M elements where N is the number of bits in each symbol. In binary modulation where $M = 2$, one bit at a time is transmitted.

2.2.2.1 M-ary Amplitude Shift Keying (M-ASK)

In ASK the amplitude of the carrier signal varies in accordance with the amplitude of the digital message signal. The concept is the same as for OOK. In M-ary ASK the amplitude of the carrier signal can take M possible levels. M-ary ASK can be represented as in Equation 2.15 [3].

$$s_i(t) = \sqrt{\frac{2E_i(t)}{T}} cos(w_0 t + \phi) \tag{2.15}$$

where $0 \leq t \leq T$, $i = 1, ..., M$, $E_i(t)$ is the energy of the symbol at i^{th} level, $\sqrt{\frac{2E_i(t)}{T}}$ represents the amplitude and can take M different values, T is the symbol duration, w_0 is the angular frequency and ϕ is a constant phase.

In this thesis, 4-ASK and 8-ASK modulated signals are considered. According to Equation 2.14, there are 2 bits per symbol for 4-ASK and 3 bits per symbol for 8-ASK. The constellation diagram for 4-ASK is given in Figure 2.4 to set an example.

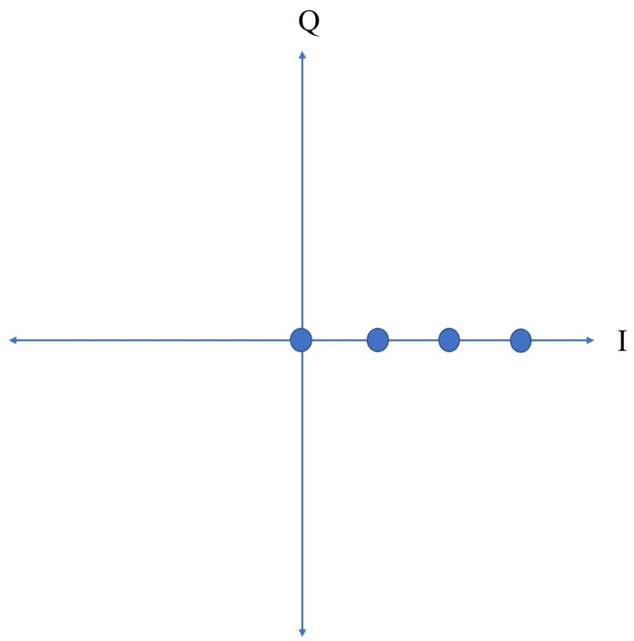

Figure 2.4: Constellation Diagram of 4-ASK

For ASK signals only amplitude variation is observed in constellations.

2.2.2.2 M-ary Phase Shift Keying (M-PSK)

In M-ary PSK, the phase of the carrier waveform is shifted by the amplitude of the modulating message signal. M-PSK modulated signal is given in Equation 2.16

$$s_i(t) = \sqrt{\frac{2E}{T}} cos(w_0 t + \phi_i(t)) \tag{2.16}$$

where $0 \leq t \leq T$, $i = 1, ..., M$, E is the energy of the symbol, $\sqrt{\frac{2E}{T}}$ represents the amplitude, T is the symbol duration, w_0 is the angular frequency and $\phi_i(t)$ is the phase which takes M different values.

In this thesis, M-PSK modulated signals for $M = 2, 4, 8, 16$ are considered. For $M = 2$ it is called Binary Phase Shift Keying (BPSK) and for $M = 4$ it is called Quadrature Phase Shift Keying (QPSK). For BPSK, the modulated signal can take two different phase values, 0 or π.

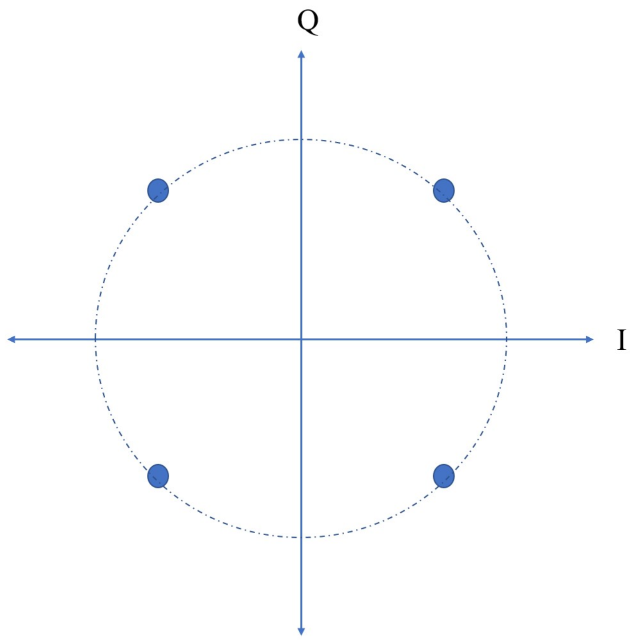

Figure 2.5: Constellation Diagram of QPSK

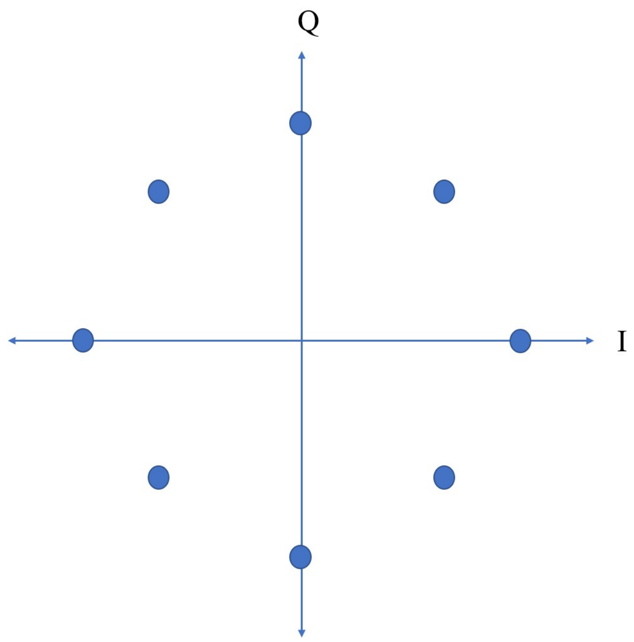

Figure 2.6: Constellation Diagram of 8-PSK

As it can be seen from Figure 2.5 and Figure 2.6, the phase component has four possible values as $\frac{\pi}{4}, \frac{3\pi}{4}, \frac{5\pi}{4}, \frac{7\pi}{4}$ for QPSK and two times more values for 8-PSK. In addition, for PSK signals only phase change is observed in constellations.

2.2.2.3 M-ary Amplitude-Phase Shift Keying (M-APSK)

APSK can be considered as the combination of ASK and PSK modulations. Both the amplitude and phase of the carrier signal is modulated by the information signal. The general expression for APSK is given in Equation 2.17;

$$s_i(t) = \sqrt{\frac{2E_i(t)}{T}} cos(w_0 t + \phi_i(t)) \tag{2.17}$$

where $0 \leq t \leq T$, $i = 1, ..., M$, E is the energy of the symbol, $\sqrt{\frac{2E_i(t)}{T}}$ represents the amplitude at i^{th} level, T is the symbol duration, w_0 is the angular frequency and $\phi_i(t)$ is the phase at i^{th} level which can take M different values.

The constellation diagram of 16-APSK is illustrated in Figure 2.7.

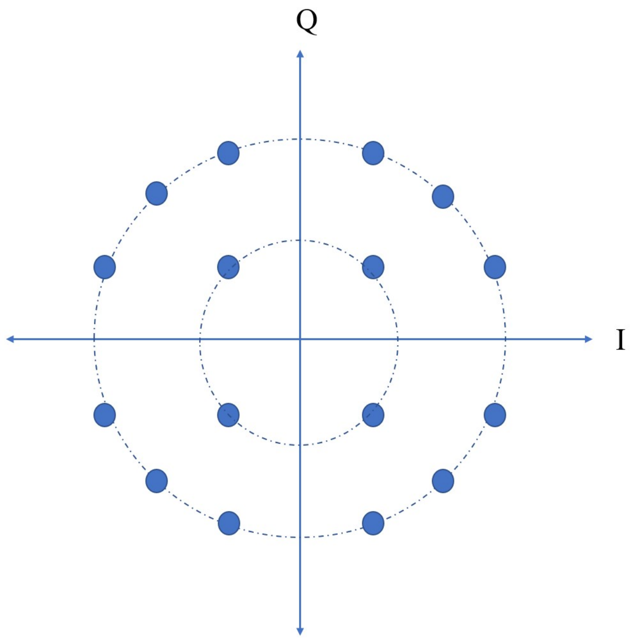

Figure 2.7: Constellation Diagram of 16-APSK

Here, four constellation points at the inner circle have the same amplitude value, and the other twelve constellation points have another same amplitude. The phase component also changes counterclockwise.

2.3 Feature Extraction

In order to create a feature space on which the classifier can operate, specific features are extracted from the data to be classified. Rather than giving raw data itself as input to the classifier, giving the features that represent the data reduces the dimension and correspondingly the complexity. This also increases the noise immunity of the classifier.

In this thesis, 41 different features are used to classify incoming signals according to modulation types. As features, higher-order statistics (HOS) i.e. higher-order moments (HOMs) and higher-order cumulants (HOCs) up to 8^{th} order are used as well as mean, variance, skewness and kurtosis values of the received signal's centered instantaneous amplitude $\alpha_c(n)$, phase $\phi_{cnl}(n)$ and frequency $f_c(n)$. Each feature is calculated on each 1024 sample of the signal.

Especially HOMs and HOCs are effective in classifying signals modulated by digital modulation techniques using symbol properties [50]. Each HOM and HOC values are unique for each modulation type. They can work properly at low SNR values due to their high noise immunity. In particular, for the analog modulation types, it is seen that the use of $\alpha_c(n)$, $\phi_{cnl}(n)$ and $f_c(n)$ in feature extraction is an effective method though can not perform well in low SNR. [51].

2.3.1 Statistical Features

Using centered signal characteristics $a_c(n)$, $f_c(n)$ and $\phi_{cnl}(n)$ given in 2.5, 2.6 and 2.8 as features directly is not preferred because it is heavy in processing load and takes up more space in dimension. Instead, by using statistics of signal characteristics i.e. mean, variance, skewness and kurtosis, the size of the feature space is reduced, and therefore the complexity. From these statistics, the mean value is used to determine the central tendency, the variance is used to identify the measure of spread, skewness and kurtosis define the shape of the distribution.

The mean value is the sum of the observation results divided by the number of observations. It gives the average value of the distribution of data. In a probabilistic distribution, the mean value can be expressed as expected value (E) in terms of probability-weighted average. In terms of moments, the mean value can be named as the first moment.

Standard deviation is calculated by dividing the sum of deviations from the mean of each observation value by the number of observations and taking the square root. Then, the variance is obtained by squaring the standard deviation. In a statistical distribution, the variance can be defined as the expectation of the square of the deviation of a random variable from the mean value. This gives information about the spread around the average value. Variance is also defined as the second central moment which is the moment about the mean.

Skewness gives a measure of the asymmetry of the statistical distribution of a random variable. Skewness refers to the distortion of the data distribution from the normal to a skewed shape to the right or left. It can be positive, negative or undefined. If the tail is on the left side of the distribution, a negative skew is observed. For the positive skew, it will be vice versa. In probability, skewness can be defined as the third standardized moment which is normalized to a representation of standard deviation. It can be calculated by dividing the third central moment by the cube of the standard deviation.

Kurtosis is a measure of how steep or flat the normal distribution curve is. It gives information about the tailedness of the probability distribution. If the kurtosis coefficient is positive, the curve is steeper than normal. The high level of kurtosis can be interpreted as having frequently excessive outliers. Negative is more flattened than normal. It is also known as the fourth standardized moment. It can be calculated by dividing the fourth central moment by the fourth power of the standard deviation. Skewness and kurtosis values indicate whether the data shows the normal distribution. In a normal distribution, the coefficients of skewness and kurtosis are zero.

Mean μ, variance σ^2, skewness γ and kurtosis κ values extracted from the $\alpha_c(n)$, $\phi_{cnl}(n)$ and $f_c(n)$ are calculated according to the equations given below;

$$\mu_x = \frac{1}{N_x}\sum_{n=1}^{N_x} x(n) \tag{2.18}$$

$$\sigma_x^2 = \frac{1}{N_x}\sum_{n=1}^{N_x}[x(n) - \overline{x}]^2 \tag{2.19}$$

$$\gamma_x = \frac{1}{\sigma_x^3 N_x}\sum_{n=1}^{N_x}[x(n) - \overline{x}]^3 \tag{2.20}$$

$$\kappa = \frac{1}{\sigma_x^4 N_x}\sum_{n=1}^{N_x}[x(n) - \overline{x}]^4 \tag{2.21}$$

here $\overline{x}$ is the mean of $x(n)$ and $x(n)$ represents $a_c(n)$, $f_c(n)$ and $\phi_{cnl}(n)$ to calculate μ_a, σ_a^2, γ_a, κ_a, μ_f, σ_f^2, γ_f, κ_f, μ_p, σ_p^2, γ_p and κ_p. Thus the features μ, σ^2, γ and κ are obtained for each $a_c(n)$, $f_c(n)$ and $\phi_{cnl}(n)$. The list of these statistical features is given in Table 2.1.

Table 2.1: Statictical Features

μ_a	Mean value of the instantaneous centered amplitude of the complex signal
μ_f	Mean value of the instantaneous centered frequency of the complex signal
μ_p	Mean value of the instantaneous centered non-linear phase of the complex signal
σ_a^2	Variance of the instantaneous centered amplitude of the complex signal
σ_f^2	Variance of the instantaneous centered frequency of the complex signal
σ_p^2	Variance of the instantaneous centered non-linear phase of the complex signal
γ_a	Skewness of the instantaneous centered amplitude of the complex signal
γ_f	Skewness of the instantaneous centered frequency of the complex signal
γ_p	Skewness of the instantaneous centered non-linear phase of the complex signal
κ_a	Kurtosis of the instantaneous centered amplitude of the complex signal
κ_f	Kurtosis of the instantaneous centered frequency of the complex signal
κ_p	Kurtosis of the instantaneous centered non-linear phase of the complex signal

2.3.2 Higher-Order Moments

In addition to the statistical features mentioned above, HOS i.e. HOMs and HOCs (from 4^{th} to 8^{th} order) are used as features especially for classifying higher-order digital modulations. The moments of random variables can be used to determine the distribution of these variables. As mentioned above, in a probability distribution, the zeroth moment gives the total probability, the first moment gives the expected value i.e. mean, the second central moment gives the variance, skewness is the third standardized moment, and the kurtosis is the fourth standardized moment which can be represented as μ_{42}. The general expression used to obtain HOMs in terms of expected value is given in Equation 2.22;

$$M_{pq} = E[x^{p-q}(x^*)^q] \tag{2.22}$$

where x is the received signal, p is the order of the moment, x^* is the complex conjugate of the received signal x and q is the power of the conjugate. In short, the expected value of the p^{th} power of x is the p^{th} moment of x. The list of moments used in this thesis is given in Table 2.2.

Table 2.2: Higher-Order Moments

2^{nd} Order Moments	M_{20}, M_{21}
4^{th} Order Moments	M_{40}, M_{41}, M_{42}, M_{43}
6^{th} Order Moments	M_{60}, M_{61}, M_{62}, M_{63}
8^{th} Order Moments	M_{80}, M_{81}, M_{82}, M_{83}, M_{84}

Based on Equation 2.22, a discrete complex received signal $r[n]$ with a total of N samples can be expressed mathematically as in Equation 2.23.

$$M_{pq} = \frac{1}{N}\sum_{n=1}^{N} r[n]^{p-q}.r[n]^{*q} \tag{2.23}$$

2.3.3 Higher-Order Cumulants

HOCs are obtained as functions of HOMs. HOCs can be calculated using expected values or moments. The HOCs and their mathematical expressions used in this thesis in terms of expected values and moments are given in Table 2.3 and Table 2.4 respectively [37].

Since $C_{20} = M_{20}$ and $C_{21} = M_{21}$, the mean value will be equal to the first cumulant and the variance will be equal to the second cumulant.

Computed HOCs are then normalized by increasing to their $\frac{2}{p}$ power due to prevent deviations at the classification step. HOCs are normalized and rescaled according to Equation 2.24;

$$\hat{C}_{pq} = C_{pq}^{\frac{2}{p}} \tag{2.24}$$

For instance, normalization of C_{62} is shown in Equation 2.25;

$$\hat{C}_{62} = C_{62}^{\frac{2}{6}} = C_{62}^{\frac{1}{3}} \tag{2.25}$$

Among the obtained HOMs and HOCs, magnitude values of the complex-valued ones are used in order to provide ease of operation and time saving while performing training in classifiers. In this way, phase shift immunity in constellation also increases [52].

Derived features mentioned in the subsections 2.3.1, 2.3.2 and 2.3.3 are given as input to ML algorithms for the classification of signals in terms of their modulation types. In order to perform the classification, a feature space is created from the derived features.

Table 2.3: Higher-Order Cumulants (HOCs) in terms of Expected Value

$C_{s,2,0}$	$E_{s,2,0} = E[x^2.(\overline{x})^0]$
$C_{s,2,1}$	$E_{s,2,1} = E[x^1.(\overline{x})^1]$
$C_{s,4,0}$	$E_{s,4,0} - 3.(E_{s,2,0})^2$
$C_{s,4,1}$	$E_{s,4,1} - 3.E_{s,2,0}.E_{s,2,1}$
$C_{s,4,2}$	$E_{s,4,2} - (E_{s,2,0})^2 - 2(E_{s,2,1})^2$
$C_{s,6,0}$	$E_{s,6,0} - 15E_{s,2,0}E_{s,4,0} + 30(E_{s,2,0})^3$
$C_{s,6,1}$	$E_{s,6,1} - 10E_{s,2,0}E_{s,4,1} - 5E_{s,2,1}E_{s,4,0} + 30(E_{s,2,0})^2E_{s,2,1}$
$C_{s,6,2}$	$E_{s,6,2} - E_{s,2,0}E_{s,4,0} - 8E_{s,2,1}E_{s,4,1} - 6E_{s,2,0}E_{s,4,2} + 6(E_{s,2,0})^3 + 24(E_{s,2,1})^2E_{s,2,0}$
$C_{s,6,3}$	$E_{s,6,3} - 6E_{s,2,0}E_{s,4,1} - 9E_{s,2,1}E_{s,4,2} + 18(E_{s,2,0})^2E_{s,2,1} + 12(E_{s,2,1})^3$
$C_{s,8,0}$	$E_{s,8,0} - 35(E_{s,4,0})^2 - 630(E_{s,2,0})^4 + 420(E_{s,2,0})^2(E_{s,4,0})$
$C_{s,8,1}$	$E_{s,8,1} - 35E_{s,4,0}E_{s,4,1} - 630(E_{s,2,0})^3E_{s,2,1} + 210E_{s,4,0}E_{s,2,0}E_{s,2,1} + 210E_{s,2,1}E_{s,4,1}$
$C_{s,8,2}$	$E_{s,8,2} - 15E_{s,4,0}E_{s,4,2} - 20(E_{s,4,1})^2 + 30E_{s,4,0}(E_{s,2,0})^2 + 60E_{s,4,0}(E_{s,2,1})^2 + 240E_{s,4,1}E_{s,2,1}E_{s,2,0} + 90E_{s,4,2}(E_{s,2,0})^2 - 90(E_{s,2,0})^4 - 540(E_{s,2,0})^2(E_{s,2,1})^2$
$C_{s,8,3}$	$E_{s,8,3} - 5E_{s,4,0}E_{s,4,1} - 30E_{s,4,1}E_{s,4,2} + 90E_{s,4,1}(E_{s,2,0})^2 + 120E_{s,4,1}(E_{s,2,1})^2 + 180E_{s,4,2}E_{s,2,1}E_{s,2,0} + 30E_{s,4,0}E_{s,2,0}E_{s,2,1} - 270(E_{s,2,0})^3E_{s,2,1} - 360(E_{s,2,1})^3E_{s,2,0}$
$C_{s,8,4}$	$E_{s,8,4} - (E_{s,4,0})^2 - 18(E_{s,4,2})^2 - 16(E_{s,4,1})^2 - 54(E_{s,2,0})^4 - 144(E_{s,2,1})^4 - 432(E_{s,2,0})^2(E_{s,2,1})^2 + 12E_{s,4,0}(E_{s,2,0})^2 + 96E_{s,4,1}E_{s,2,1}E_{s,2,0} + 144E_{s,4,2}(E_{s,2,1})^2 + 72(E_{s,4,2})(E_{s,2,0})^2 + 96E_{s,4,1}E_{s,2,0}E_{s,2,1}$

Table 2.4: Higher-Order Cumulants (HOCs) in terms of Moments

C_{20}	M_{20}
C_{21}	M_{21}
C_{40}	$M_{40} - 3M_{20}^2$
C_{41}	$M_{41} - 3M_{20}M_{21}$
C_{42}	$M_{42} - M_{20}^2 - 2M_{21}^2$
C_{60}	$M_{60} - 15M_{20}M_{40} + 30M_{20}^3$
C_{61}	$M_{61} - 10M_{20}M_{41} - 5M_{21}M_{40} + 30M_{20}^2M_{21}$
C_{62}	$M_{62} - M_{20}M_{40} - 8M_{21}M_{41} - 6M_{20}M_{42} + 6M_{20}^3 + 24M_{21}^2M_{20}$
C_{63}	$M_{63} - 6M_{20}M_{21} - 9M_{21}M_{42} + 18M_{20}^2M_{21} + 12M_{21}^3$
C_{80}	$M_{80} - 35M_{40}^2 - 630M_{20}^2 + 420M_{20}^2M_{40}$
C_{81}	$M_{81} - 35M_{40}M_{41} - 630M_{20}^3M_{21} + 210M_{40}M_{20}M_{21} + 210M_{20}M_{41}$
C_{82}	$M_{82} - 15M_{40}M_{42} - 20M_{41}^2 + 30M_{40}M_{20}^2 + 60M_{40}M_{21}^2 + 240M_{41}M_{21}M_{20} + 90M_{42}M_{20}^2 - 90M_{20}^4 + 540M_{20}^2M_{21}^2$
C_{83}	$M_{83} - 5M_{40}M_{41} - 30M_{41}M_{42} + 90M_{41}M_{42} + 120M41M_{21}^2 + 180M_{42}M_{21}M_{20} + 30M_{40}M_{20}M_{21} - 270M_{20}^3M_{21} - 360M_{21}^3M_{20}$
C_{84}	$M_{84} + M_{40}^2 - 18M_{42}^2 - 16M_{41}^2 - 54M_{20}^4 - 144M_{21}^4 - 432M_{20}^2M_{21}^2 + 12M_{40}M_{20}^2 + 966M_{41}M_{21}M_{20} + 144M_{42}M_{21}^2 + 72M_{42}M_{20}^2 + 72M_{42}M_{21}^2 + 96M_{41}M_{20}M_{41}$

2.4 Classification

In order to map extracted features to a feature space and classify data using different methods, the performances of many different classifiers have been compared with each other. There is a wide range of classifier types with performances that vary depending on the features used, processing time and complexity. Due to the nature of the classifiers, their success performance increases as their complexity increases. Achieving high performances by keeping complexity as low as possible is an important criterion in classifier design.

The types of digital modulations in dataset range from low orders (binary) to high orders (256-ary). Dataset used in this thesis also offers 16 different SNR options in the range of 0 to +30 dB. Thus, high order modulated signals, which are more difficult to classify and can be confused more easily in low SNRs, can be examined in both low and high SNR levels as in the practice. High order modulation types require high SNR values for high accuracy in the real world. On the other hand, it is much easier to distinguish modulation types such as AM, FM, PM with specific features and low order modulation types such as BPSK, BFSK ($M = 2$) in low SNRs. When higher orders are reached ($M > 2$), it is observed that M-PSK and M-QAM types are more confused as expected, especially considering the modulation structures of these types are similar.

When comparing classifiers' performances, Classification Learner Toolbox of Matlab which is a machine learning application is used. With the supervised learning approach, some of the data is used to train the classifier and the rest to test the trained classifier. A class label is assigned to each data in the dataset allocated for training. The algorithm is intended to learn which type of data should be in which class and build a model. The algorithm is then given the data allocated for the test and is asked to assign a class label to each of them according to the obtained model. The success of the algorithm is measured by how much the algorithm assigns the correct class labels to the data in the test set. Classifiers available in Classification Learner Toolbox of Matlab are examined. The results showed that linear, quadratic and cubic SVM classifiers outperform among them for 41 different features.

80% of the data is used for training and 20% for the test. As the amount of data allocated for training increases, so does the performance of the classifier. The classification models that have high flexibility could model small changes in the data even they haven't any meanings such as noise so the overfitting may occur. 5-fold cross-validation is also used to prevent overfitting. The data set is divided into 5 subsets, one of which is used for validation.

SVM has a non-parametric model. In this model type, there is no certain number of parameters that constitutes the model. SVM allows binary or multiclass predictions. If the data will be divided into two different classes, this is called binary classification. If there are more than two classes of data to be allocated, this is a multiclass classification problem. The prediction speed could be slower and the usage of memory could be higher in multiclass predictions. When SVM performs multiclass classification, the algorithm reduces the problem to a set of binary classification.

The kernel function is specified so that the classifier can operate as linear, cubic or quadratic. Linear SVM which has low flexible model separates data belongs to different classes using linear boundaries. Quadratic and cubic SVM have slightly higher flexibility according to linear SVM.

Linear SVM tries to classify data in the sample space in a linear manner using hyperplanes as decision mechanisms. Suppose there is a feature space created from data belonging to two different classes. As illustrated in Figure 2.8 the distance of the hyperplane passing through the data of different classes that are closest to each other must be maximum. This gives us optimal separating hyperplane. The one with the highest margin is selected from the infinite lines that can separate the data. The data belonging to two different classes located closest to each other is called support vector [53]. The robustness increases as the best margin expand.

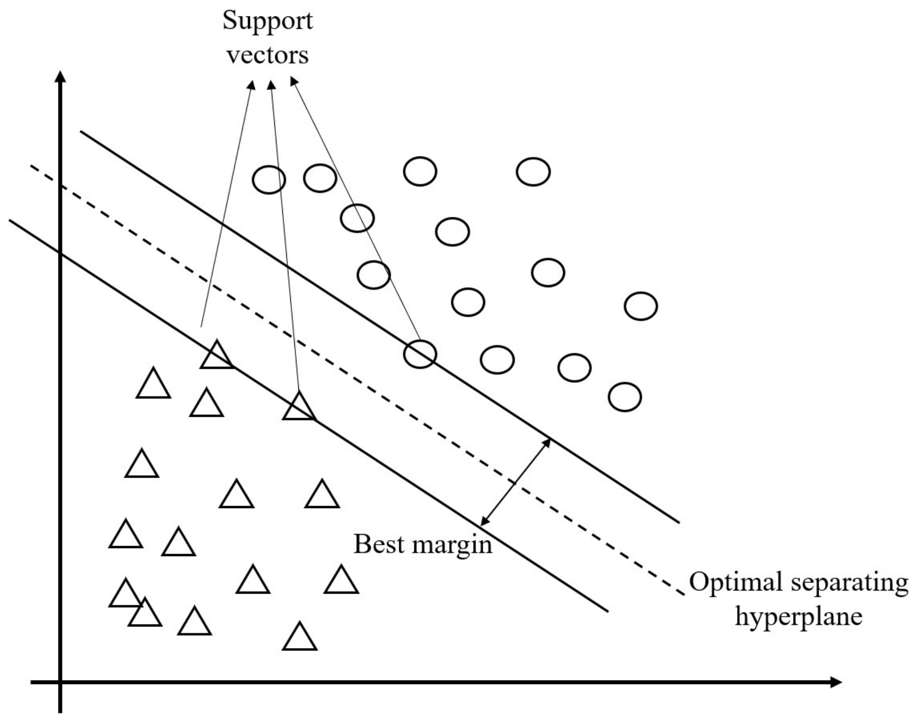

Figure 2.8: Linear SVM in two-dimensional feature space

The general formulation for SVM is given in Equation 2.26.

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{2.26}$$

where $K(x, x_i)$ is the Kernel function of the feature model (x, x_i), x_i is the input vector of the i^{th} training example, x is the input vector of the test example, N is the number of support vectors, α_i represents the i^{th} support vectors in the case of $\alpha_i \neq 0$ and b is the constant bias. The decision of the classification is made by looking at the sign of the equation.

Also, with using kernel function of the linear SVM, non-linear distinctions between classes can be conducted. This is called kernel trick and the idea is based on mapping the input features to a higher dimensional feature space and determining the best margin. Kernel function identifies the SVM as linear, quadratic and cubic which is also known as polynomial kernel. In the equation for polynomial kernel given in 2.27, the order of the kernel n is equal to 1 for linear, 2 for quandratic and 3 for cubic kernel.

$$K(x, x_i) = (xx_i + c)^n \tag{2.27}$$

where c is constant which is equal to zero for homogenous polynomial and equal to one for inhomogenous polynomial. It provides trade-offs in the effects of high and low-order terms.

The variations of hyperplanes in quadratic and cubic SVMs are shown in Figure 2.9 and 2.10 respectively.

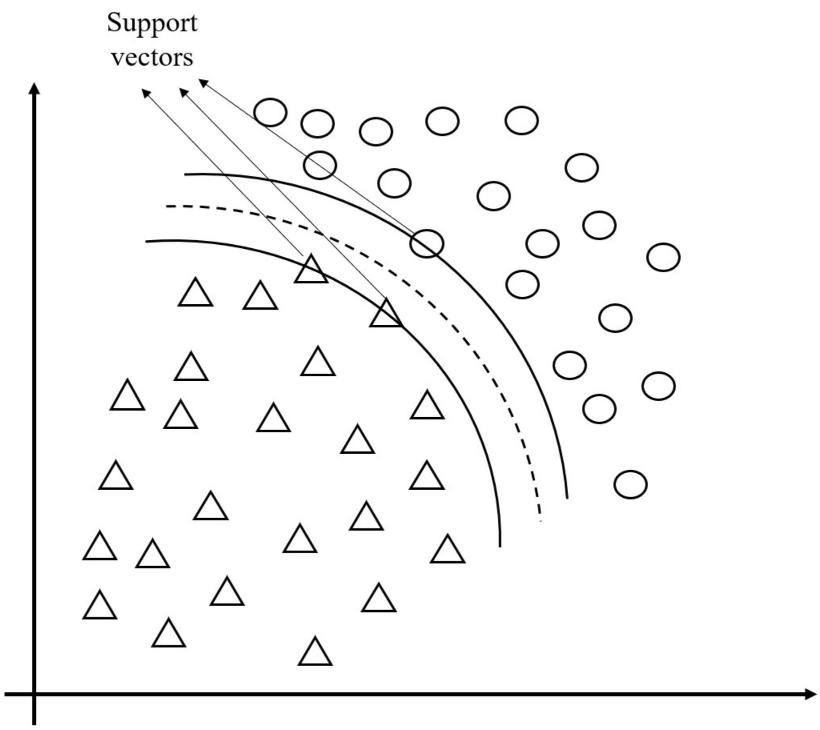

Figure 2.9: Quadratic SVM in two-dimensional feature space

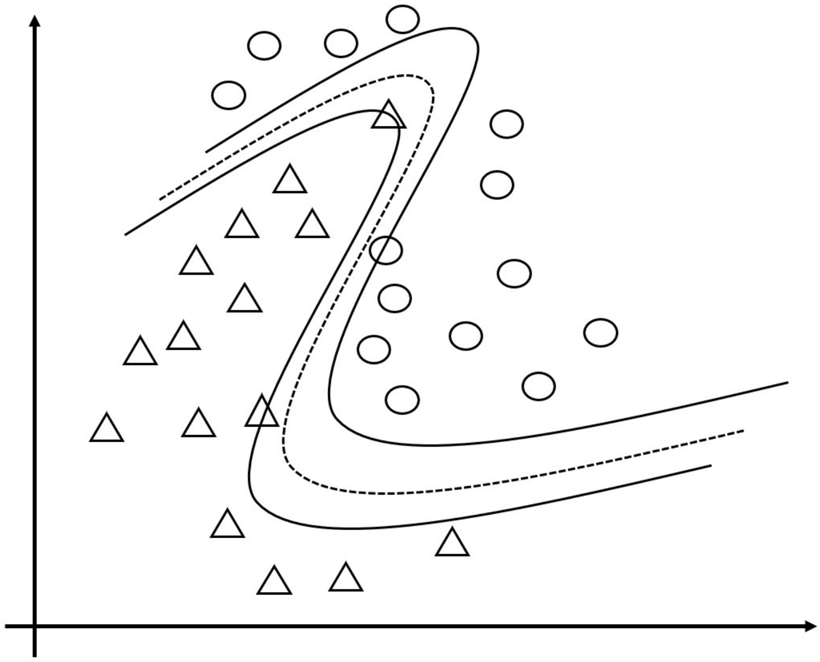

Figure 2.10: Cubic SVM in two-dimensional feature space

CHAPTER 3

RESULTS

In this chapter, after the description of the dataset and construction of the feature space, the results of the developed modulation classification (MC) algorithms by using a dataset of modulated signals with 12 different modulation techniques are given. The performances of the two different classification structures titled in this thesis as the ”traditional” and ”hierarchical” are presented.

3.1 Description of the Dataset

In this study, a dataset available in [54] is used. This dataset contains modulated radio signals collected over-the-air with the aid of software-defined radio (SDR) (USRP B210 SDR) in the laboratory environment. The use of radio signals provides a wide variety of convenient signal types for machine learning applications. Wireless data transmission, which is becoming more and more widespread nowadays, necessitates the processing of this data in different conditions and SNR levels [55].

The dataset contains radio signals modulated by several modulation types. In this thesis, signals modulated with 12 different modulation types listed in Figure 2.1 obtained from the dataset is used.

The signals from each modulation type have 16 different SNR (E_s/N_0) values that increase by two by two in the range of 0 to +30 dB. From each SNR value, there are 4096 signals (example) of each modulation type. In-phase (I) and quadrature (Q) values of each signal are sampled with 1024 samples by short time observations to meet real-life needs. The length of all samples in the dataset is the same. There are a total of $4096 \times 16 = 65536$ signals in all dB values of one modulation type. Dataset used in this thesis contains total $65536 \times 12 = 786432$ signals (examples) with 1024 I and Q samples in all SNR values of 12 different modulation types. The dataset with the dimensions of $2 \times 1024 \times 786432$ for all SNR values is obtained. For a more comprehensive explanation, the dimensions of the dataset are illustrated in Figure 3.1.

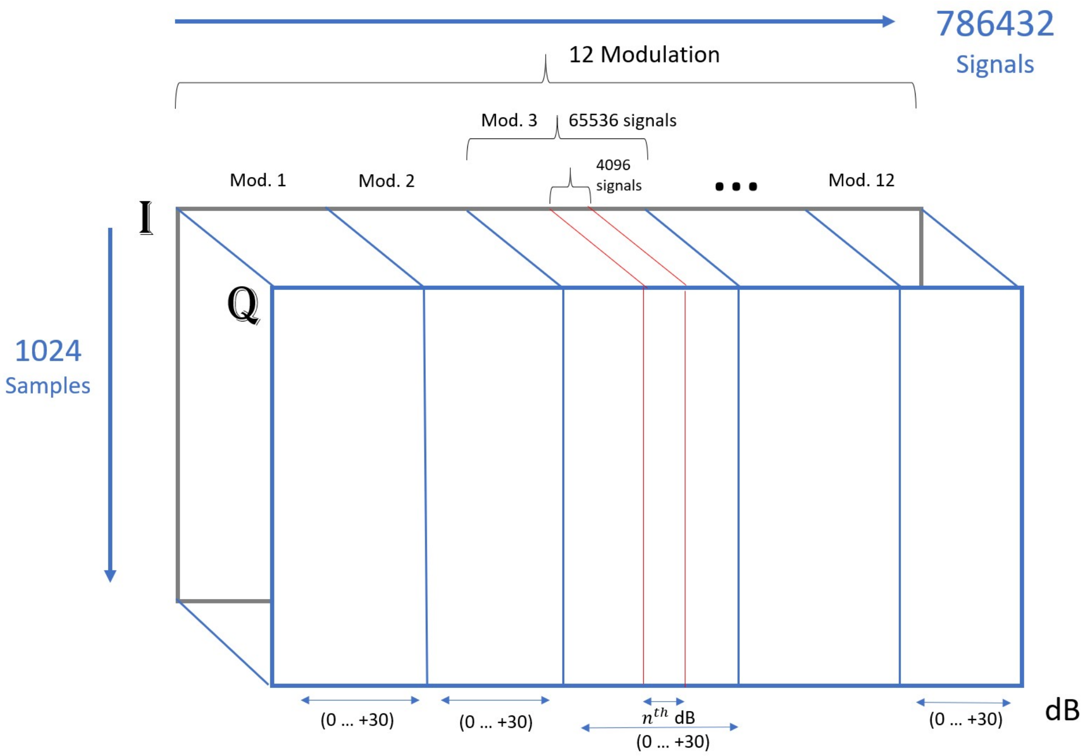

Figure 3.1: The Representation of the Dataset

3.2 Construction of the Feature Space

Each of the features used in this study given in Table 2.1, 2.2 and 2.4 is calculated over each I and Q values with 1024 samples. After this step the feature set is illustrated in Figure 3.2.

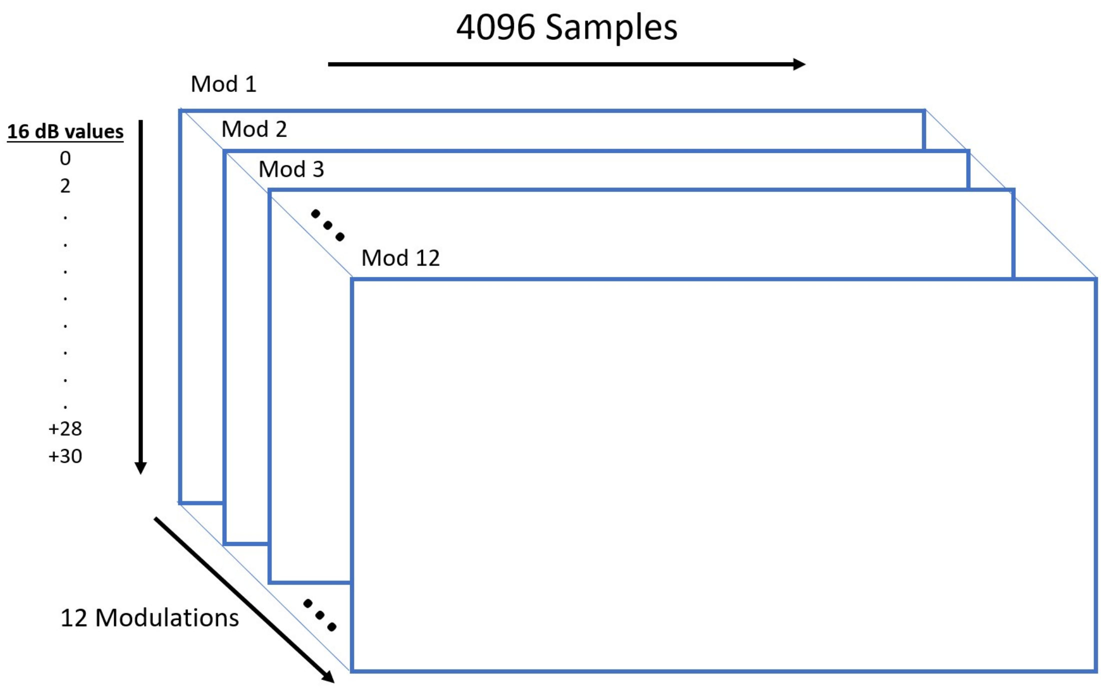

Figure 3.2: The Representation of Feature Space for One Feature

Here for a feature, we have the dimensions of $12 \times 16 \times 4096$ with 12 different modulation types, 16 different SNR values and 4096 different examples extracted from every 1024 samples. 80% is allocated for training and 20% is for testing. For training $12 \times 16 \times 3277$, for testing $12 \times 16 \times 819$ dimensions are obtained.

The feature set is then reshaped to make it suitable for input to the classifier. For one SNR value, the feature samples belong to each modulation type is line up one under other. Then all the 41 features are combined to form a class table to be an input to the classifier. A class table for one SNR value as training input is given in Figure 3.3.

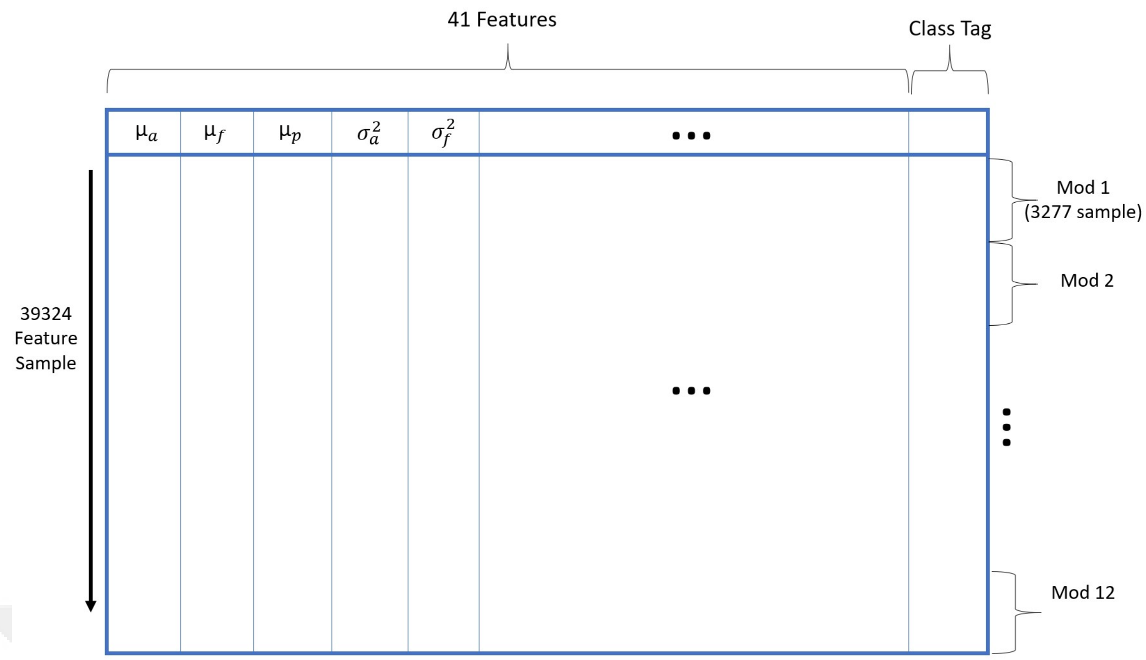

Figure 3.3: The Representation of a Feature Space at one SNR value

At first, we had the dataset with the dimensions of $2 \times 1024 \times 49152$ at one SNR value for all modulations. After the reduction of dimension for the classifier to operate effectively, the dimensions of training feature space is obtained as 39324×42. The last column is allocated for a class label which gives the information about which feature sample belongs to which modulation type. After the training process of the selected classifiers, for testing the other 20% is given to the classifier. The test set has dimensions of 9828×41 without any class label.

3.3 Classification

The classification of the modulation types is carried out using Classification Learner toolbox of Matlab. Among the classifiers available in Classificaton Learner, SVM using linear, quadratic and cubic kernel is selected which outperform according to others. First, 12 different modulation types are trained and tested using single classifier and the performances are compared. Then a hierarchical classification structure is proposed to improve the accuracy especially for higher-order modulations.

3.3.1 Traditional Structure

In the traditional structure, a single classifier is used. SVMs using linear, quadratic and cubic kernels are trained and tested with the signals modulated with total 12 modulations. Modulated signals with the modulation types shown in Figure 2.1 are classified using linear, quadratic and cubic SVMs and their performances are compared in 16 different SNR levels from 0 dB to 30 dB. The overall performance graph for the classifiers is given in Figure 3.4. The confusion matrices for quadratic SVM at 0 dB, 10 dB and 20 dB are presented in Figure 3.5, 3.6 and 3.7 respectively to set an example.

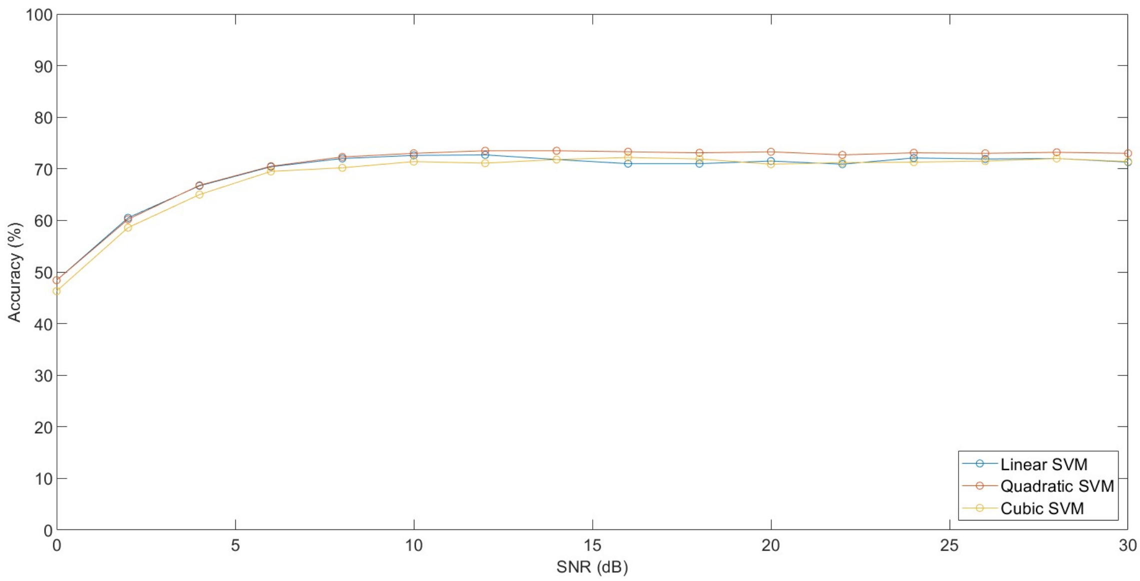

Figure 3.4: Accuracy against SNR for linear, quadratic and cubic SVMs

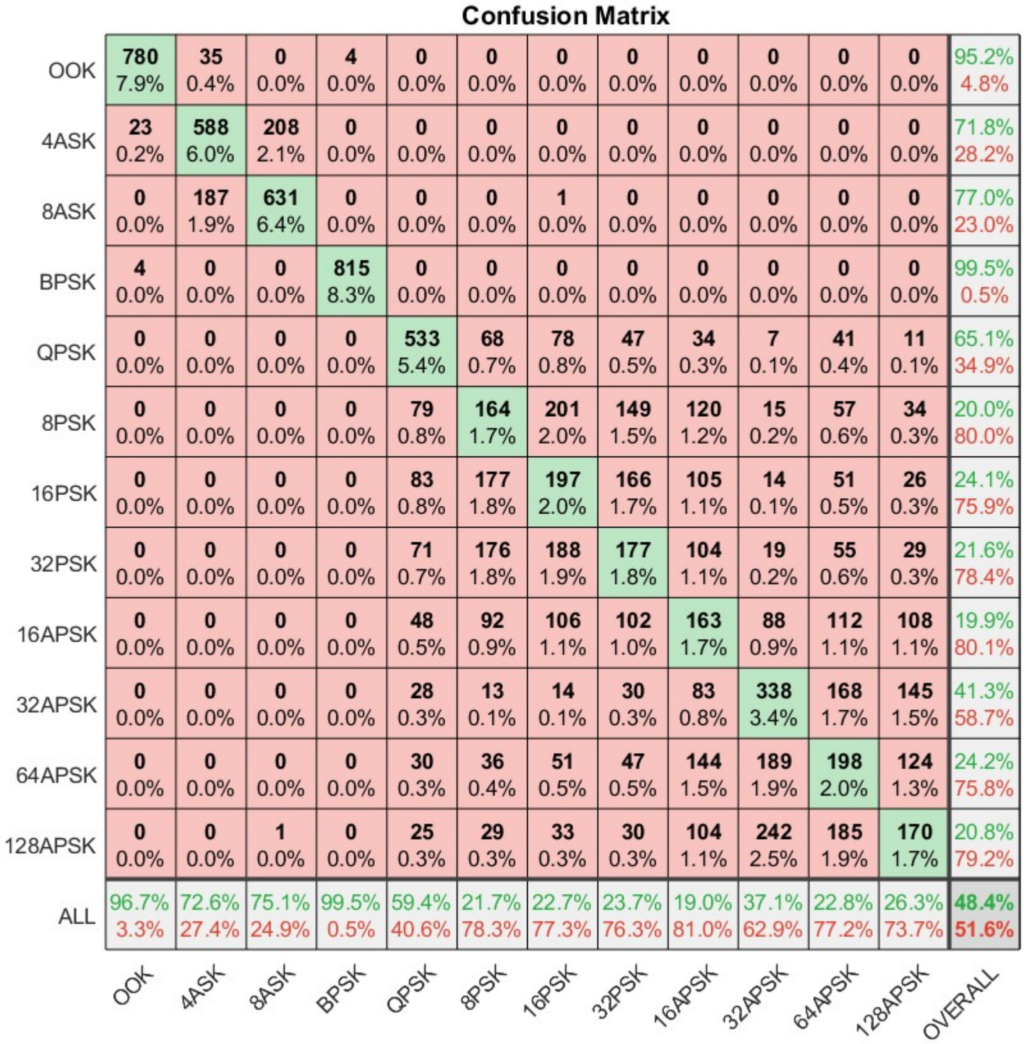

Figure 3.5: The confusion matrix of quadratic SVM for 12 modulations at 0 dB

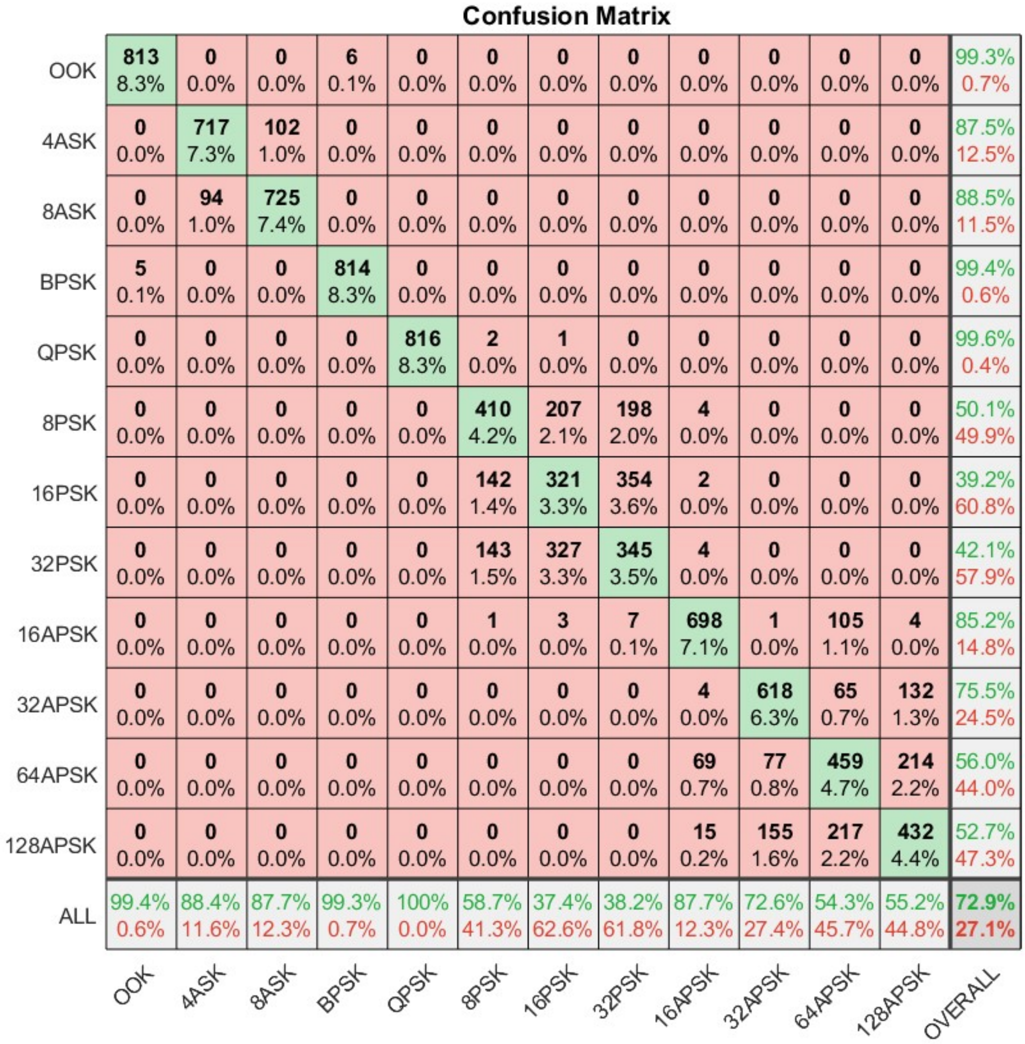

Figure 3.6: The confusion matrix of quadratic SVM for 12 modulations at 10 dB

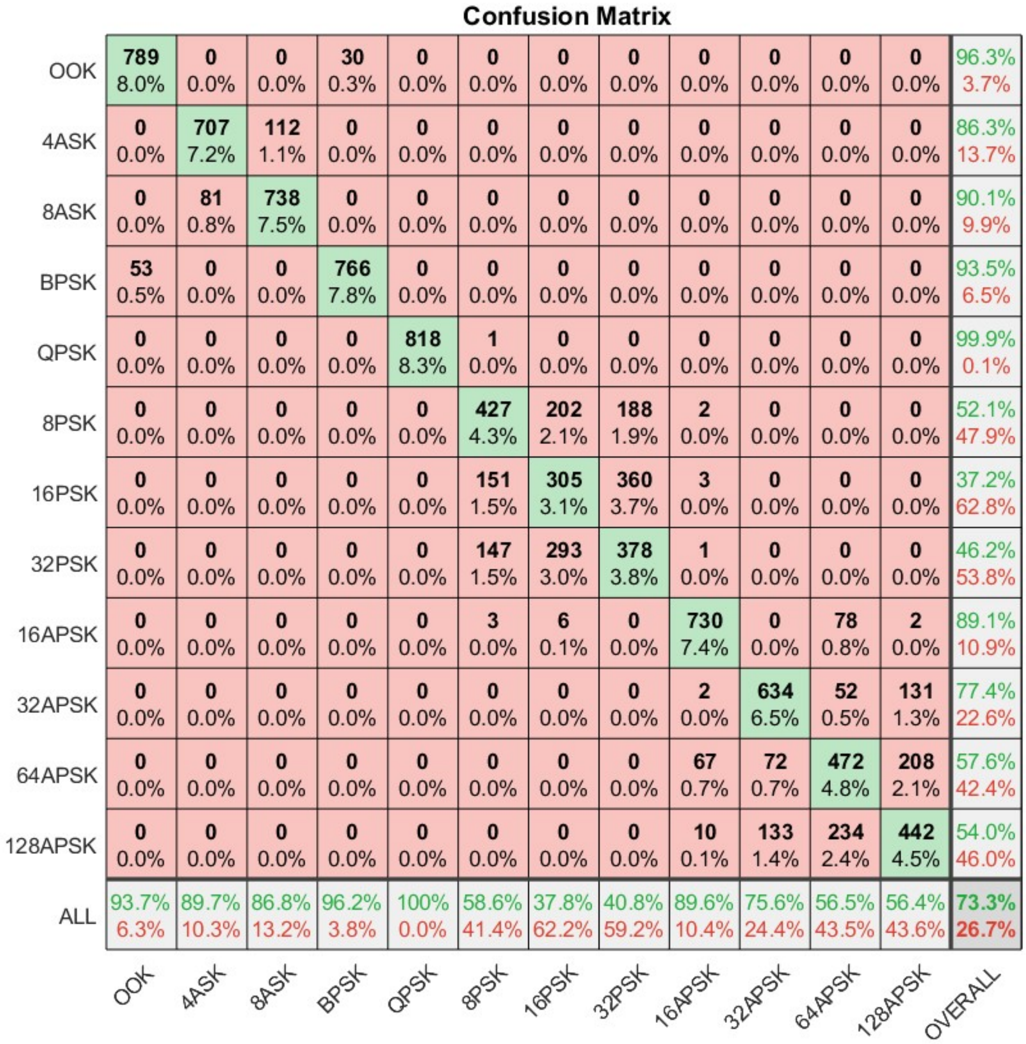

Figure 3.7: The confusion matrix of quadratic SVM for 12 modulations at 20 dB

3.3.2 Hierarchical Structure

A hierarchical structure has been established in order to classify the modulation type of a dataset containing modulated signals with an unknown modulation type with a sufficient number of samples $N \geq 100$ approximately. As N grows, the accuracy increases. By this means, in addition to OOK, 8-ASK, BPSK, QPSK and 32-APSK, the modulation types 4-ASK, 32-PSK, 16-APSK, 64-APSK, 128-APSK aims to be classified with higher performances. As classifier, quadratic SVM is chosen because of its slightly higher performance as shown in Figure 3.4. In the first step, the classifier is trained by 12 different modulations. Among those modulations, OOK, 8-ASK, BPSK, QPSK and 32-APSK are assumed to classified correctly. If the modulation type of the dataset is classified as one of these modulations, it considered correct. The performance graphs for these modulations are given in the Figures from 3.9 to 3.13. Otherwise, if the modulation type of the dataset is classified as any other 7 modulations, the dataset is then subjected to a two-stage classification to achieve a better performance compared with the traditional classifier mentioned in the previous subsection. A general overview for the hierarchical structure is given in Figure 3.8.

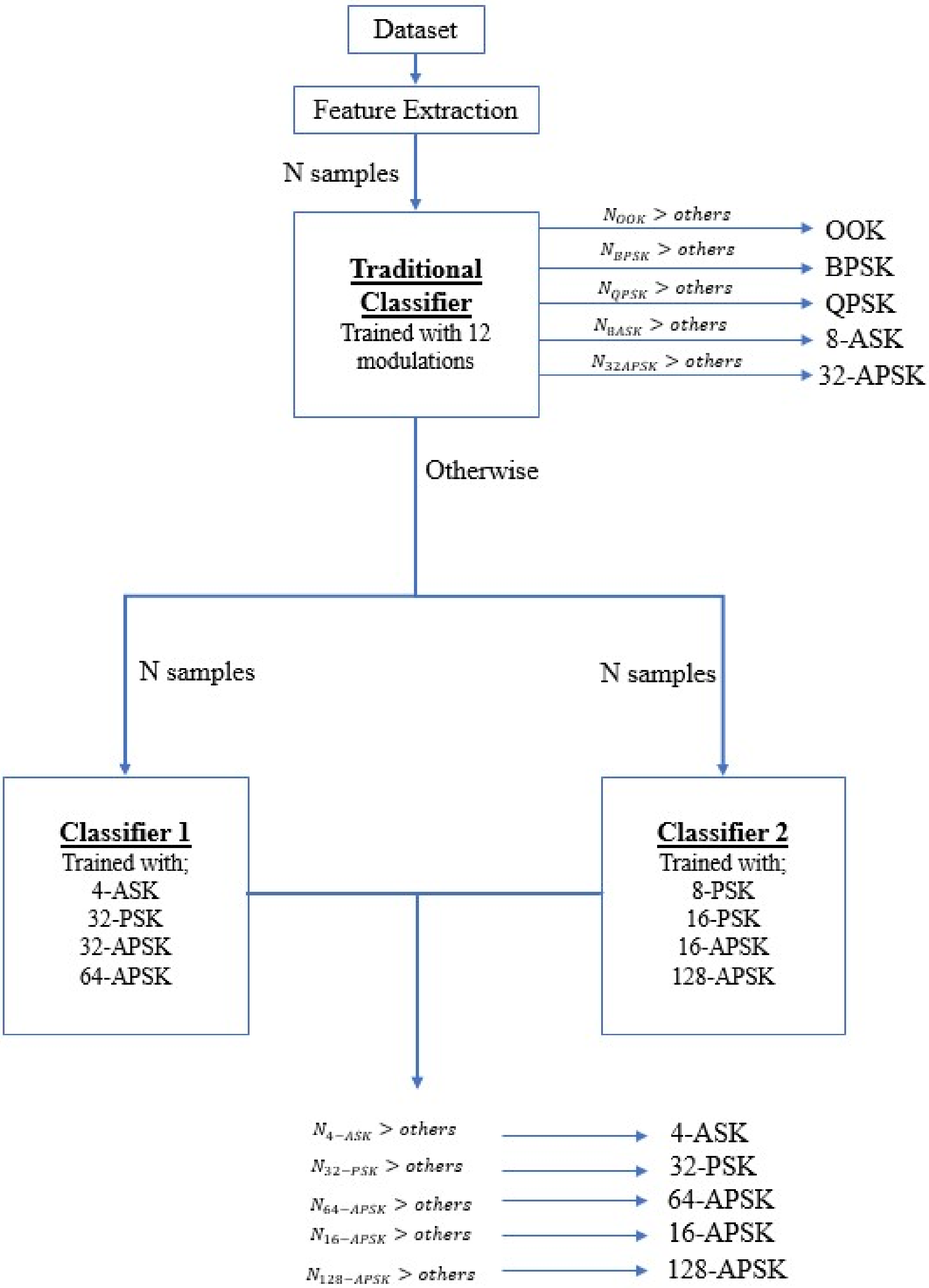

Figure 3.8: A General Overview of Hierarchical Structure

In the hierarchical structure, with which modulation types that the sub-classifiers are trained are decided according to their performance in the traditional method. The priority here is to improve the performance of high-order modulation types. As can be seen from the confusion matrices given from 3.5 to 3.7, 8-PSK, 16-PSK and 32-PSK signals are confused with each other highly. Sub-classifier 1 (C1) is trained with the 32-PSK signal, and sub-classifier 2 (C2) is trained with the 8-PSK and 16-PSK signals together to distinguish the higher-order 32-PSK signal from the others. Thus, when 32-PSK signal is given to both classifiers for testing, it can be easily distinguished in C1 while confusing between 8-PSK and 16-PSK in C2 will reduce the separation percentage. In Figure 3.14, testing accuracies for 32-PSK signal for both classifiers at 8 dB can be examined. In the comparison between the two classifiers, the ratio of C1 will be victorious and correctly classified. For this reason, however, when the 8-PSK and 16-PSK signals are given to the hierarchical structure for testing, these signals will not be successfully separated because they will interfere with each other. The trade-off here is to distinguish the higher-order modulation with high success rate. When whichever modulation order to be classified is left alone in one classifier and the others are trained in the another one, it will be successfully separated. With the same logic, both 32-APSK and 64-APSK signals are mostly confused with 128-APSK. However, they are not confused with each other huge amount. Thus, training C1 with both 32-APSK and 64-APSK and training C2 with 128-APSK provides distinguish 128-APSK with significantly high success rate. The shortcoming here is about 32-APSK signal. When 32-APSK signal is given to C1, it can be distinguish with the rate of 85%. However, when it is given to C2, it is predicted as 128-APSK with higher rate since there is no another modulation type it can confuse. For this reason 32-APSK signal is classified in the first stage. In the second stage, it is used for training to distinguish 128-APSK signal successfully.

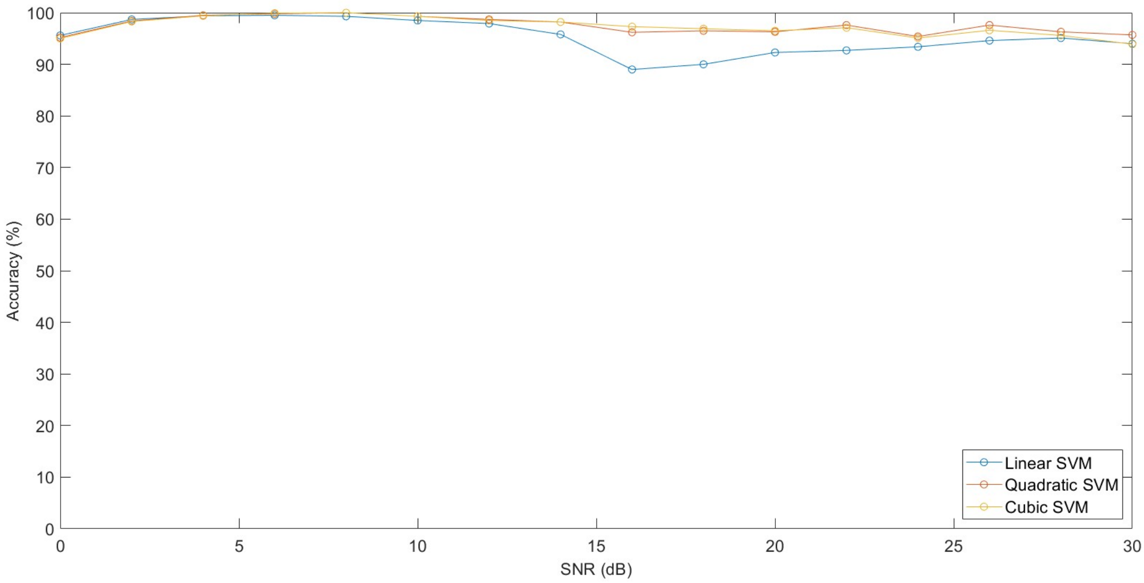

Figure 3.9: Accuracy for OOK modulation

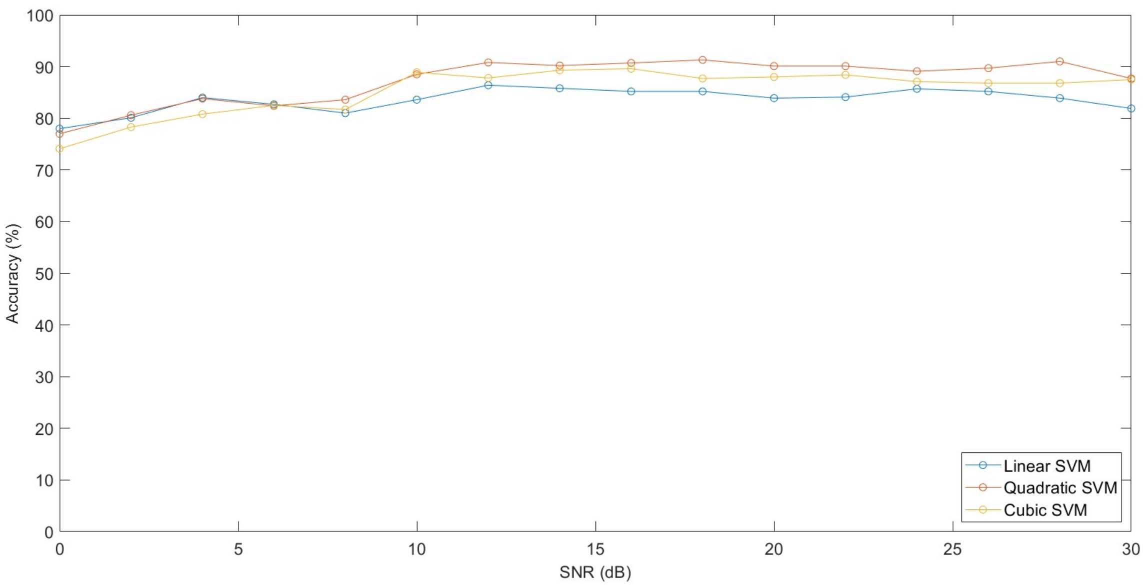

Figure 3.10: Accuracy for 8-ASK modulation

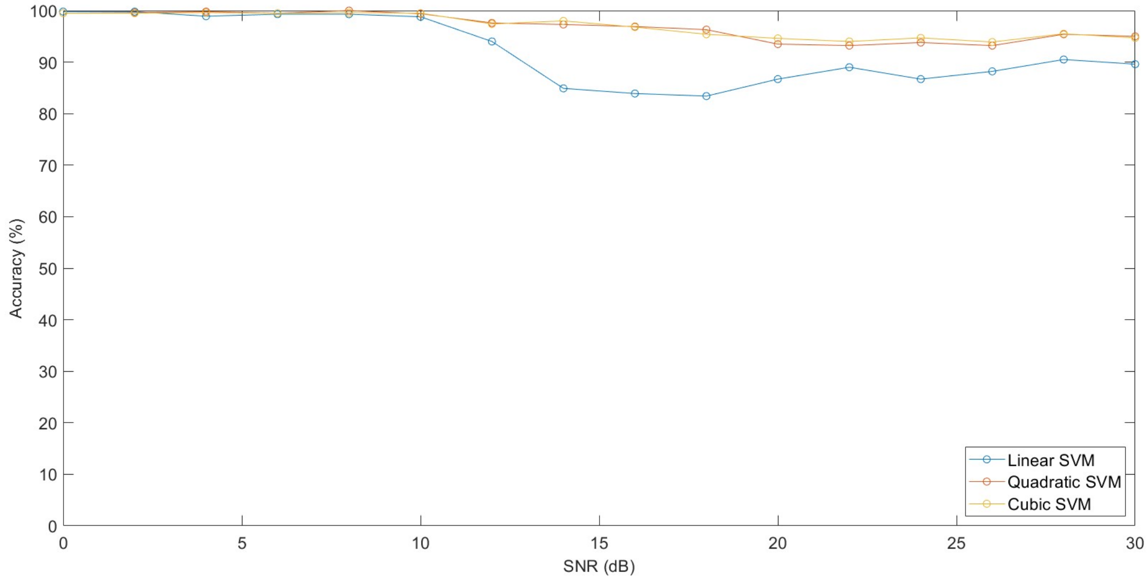

Figure 3.11: Accuracy for BPSK modulation

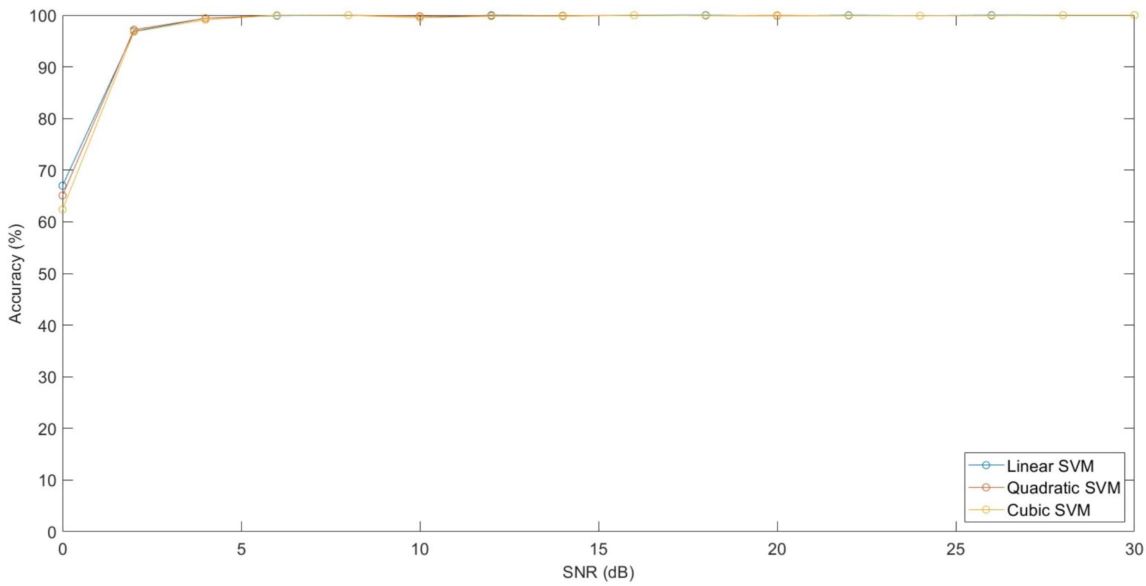

Figure 3.12: Accuracy for QPSK modulation

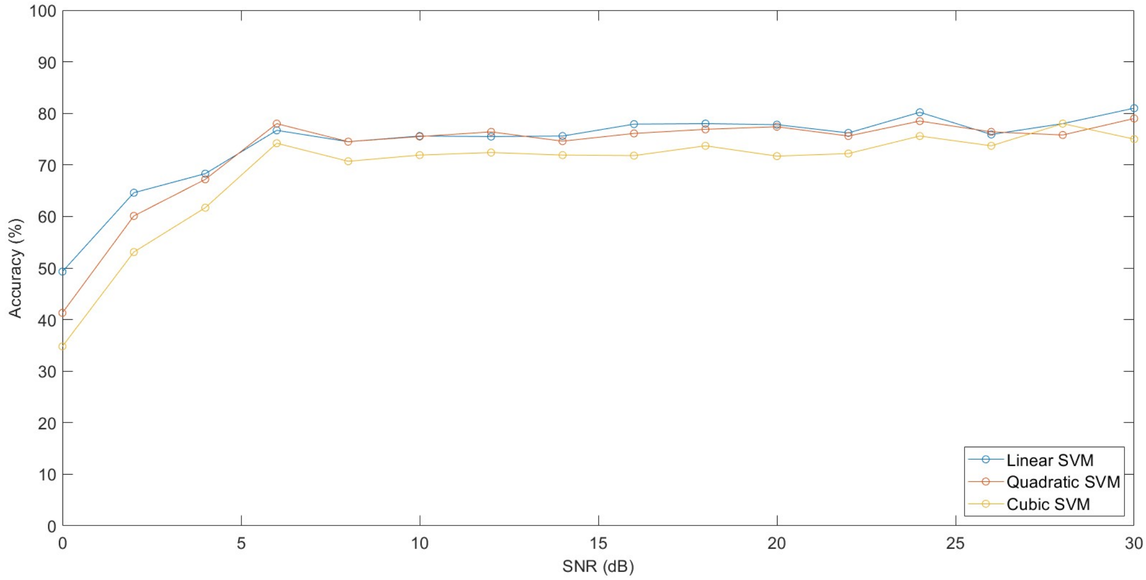

Figure 3.13: Accuracy for 32-APSK modulation

In the case of the dataset is not classified as OOK, 8-ASK, BPSK, QPSK or 32-APSK modulations in the first stage, then two different classifiers are constituted. One of the classifiers is trained with 4-ASK, 32-PSK, 32-APSK and 64-APSK while the other one is trained with 8-PSK, 16-PSK, 16-APSK, 128-APSK. The incoming dataset is tested on both of the classifiers. The modulation type separated by the largest number of samples is considered correct. In Table 3.1, modulation types and their corresponding modulation numbers in sub-classifier 1 (C1) and sub-classifier 2 (C2) is shown. Here ”MOD 1” refers to 4-ASK in C1 and it refers to 8-PSK in C2.

Table 3.1: Correspondence of modulation types in C1 and C2 sub-classifiers

	C1	C2
MOD 1	4-ASK	8-PSK
MOD 2	32-PSK	16-PSK
MOD 3	32-APSK	16-APSK
MOD 4	64-APSK	128-APSK

To set an example, in Figure 3.14 the results of the two sub-classifiers for the relevant modulation types at 8 dB SNR are presented.

		C1	C2
4-ASK	MOD 1	100	78,8
	MOD 2		
	MOD 3		10,9
	MOD 4		10,2

		C1	C2
16-APSK	MOD 1		
	MOD 2	10,8	
	MOD 3		93,7
	MOD 4	88,4	5,4

		C1	C2
32-PSK	MOD 1		25,39
	MOD 2	100	74
	MOD 3		
	MOD 4		

		C1	C2
128-APSK	MOD 1		
	MOD 2		
	MOD 3	34,6	4,7
	MOD 4	65,3	95,2

		C1	C2
64-APSK	MOD 1		
	MOD 2		
	MOD 3	13,4	24
	MOD 4	86,6	75,9

Figure 3.14: Accuracies of Modualtions for Sub-classifiers C1 and C2 at 8 dB

To make a more comprehensive explanation of the structure the following inferences can be drawn from Figure 3.14; the 32-PSK signal is given to C1 and C2 for testing. It is predicted as 32-PSK with 100% accuracy in C1 while it is predicted as 16-PSK with the accuracy of 74% and predicted as 8-PSK with the accuracy of $25,3\%$ in C2. Putting 8-PSK and 16-PSK in the same classifier for training divides the samples thus, reduces the rate of 32-PSK signal in C2. The highest rate is chosen among all the modulation types in C1 and C2 so it is classified correctly.

The performance comparison of traditional 12 modulations classifier structure and the hierarchical classifier structure for 4-ASK, 32-PSK, 16-APSK, 64-APSK and 128-APSK modulation types are given in the figures from 3.15 to 3.19.

The average accuracy for 4-ASK modulation increased from 85% to 100% as shown in Figure 3.15;

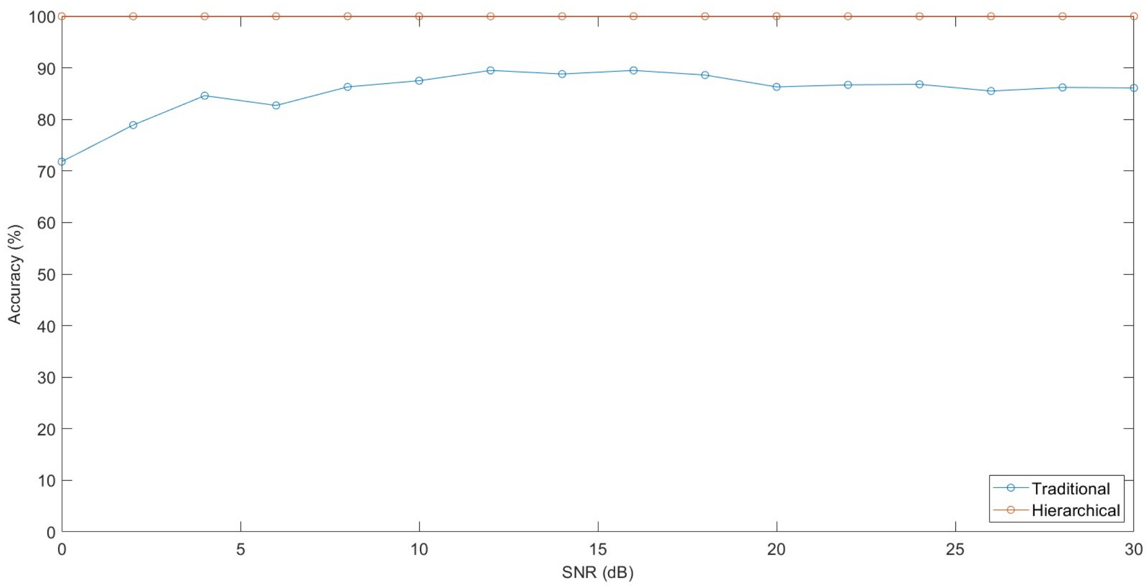

Figure 3.15: Accuracy for 4-ASK modulation

The success rate of classification for 32-PSK modulation increased from 40% to 100% as shown in Figure 3.16;

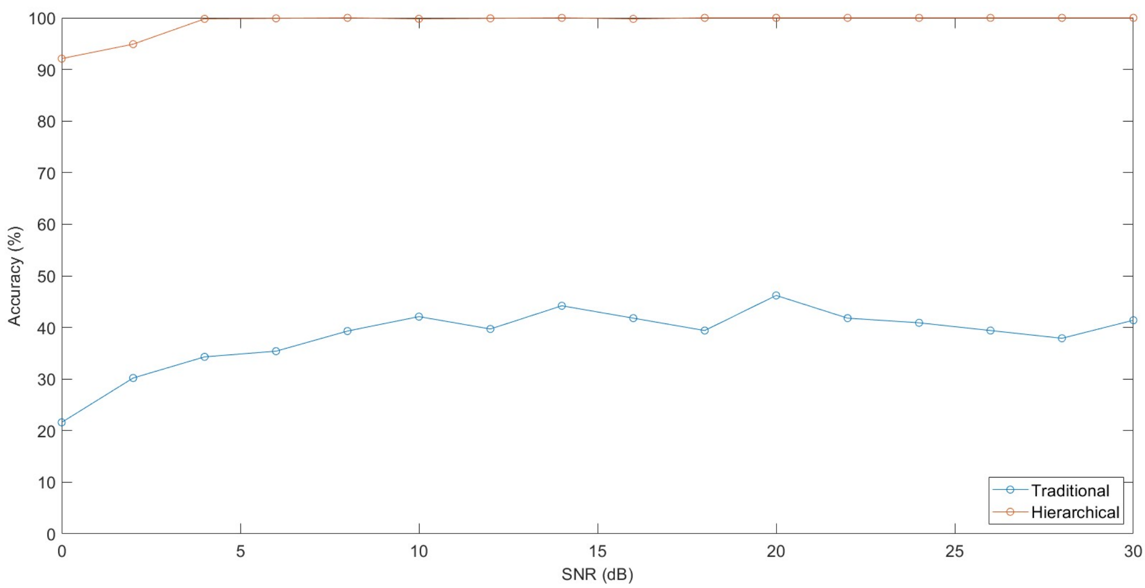

Figure 3.16: Accuracy for 32-PSK modulation

For 16-APSK modulation the accuracy is increased from 85% to 95% as shown in

Figure 3.17;

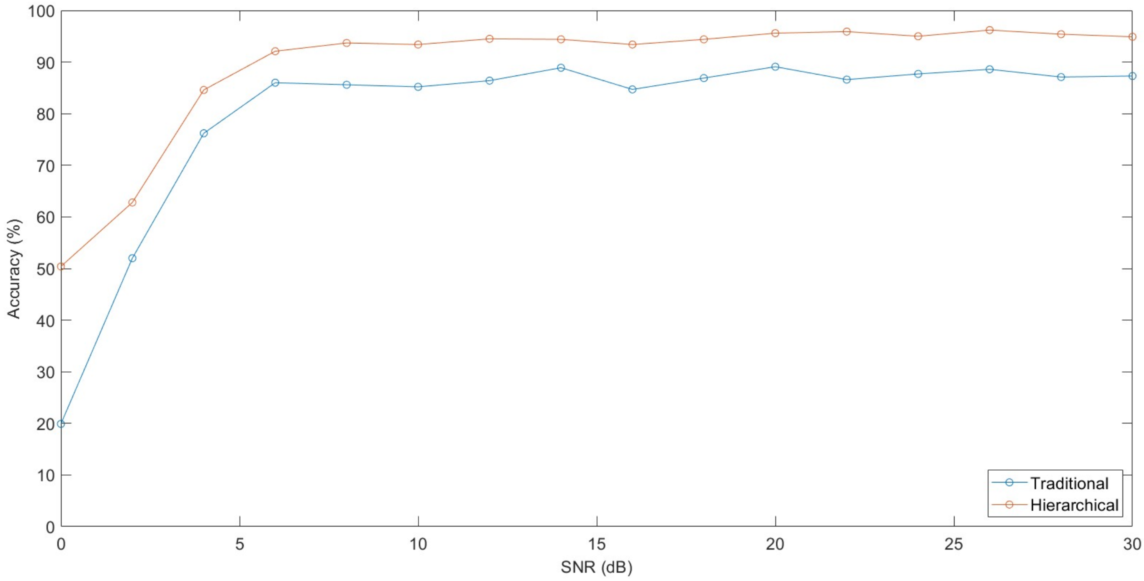

Figure 3.17: Accuracy for 16-APSK modulation

The success rate of classification for 64-APSK modulation increased from 55% to 98% as shown in Figure 3.18;

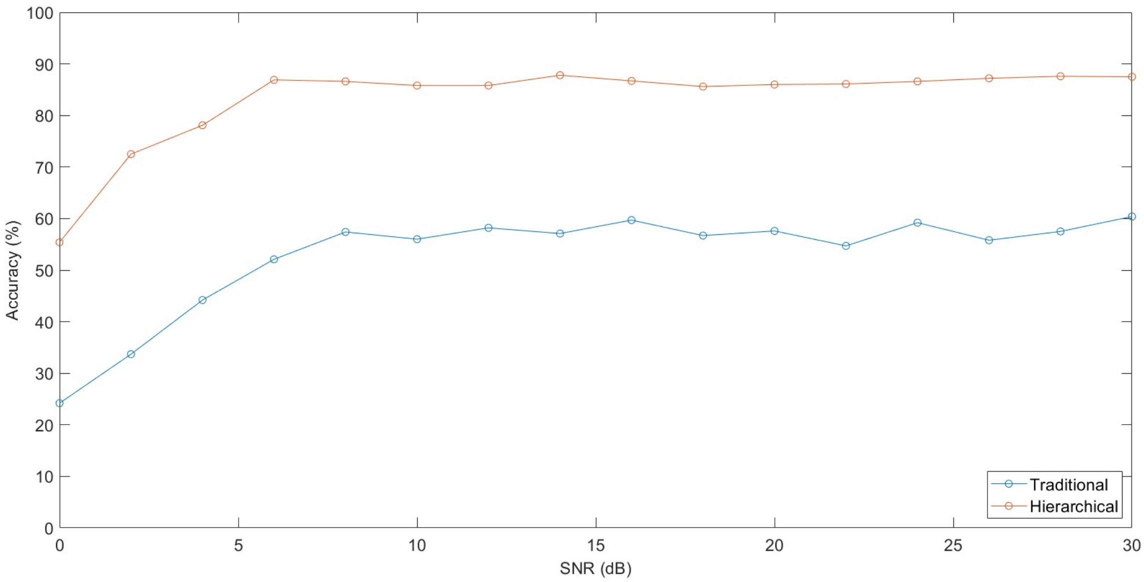

Figure 3.18: Accuracy for 64-APSK modulation

The average accuracy for 128-APSK modulation increased from 50% to 95% as shown in Figure 3.19;

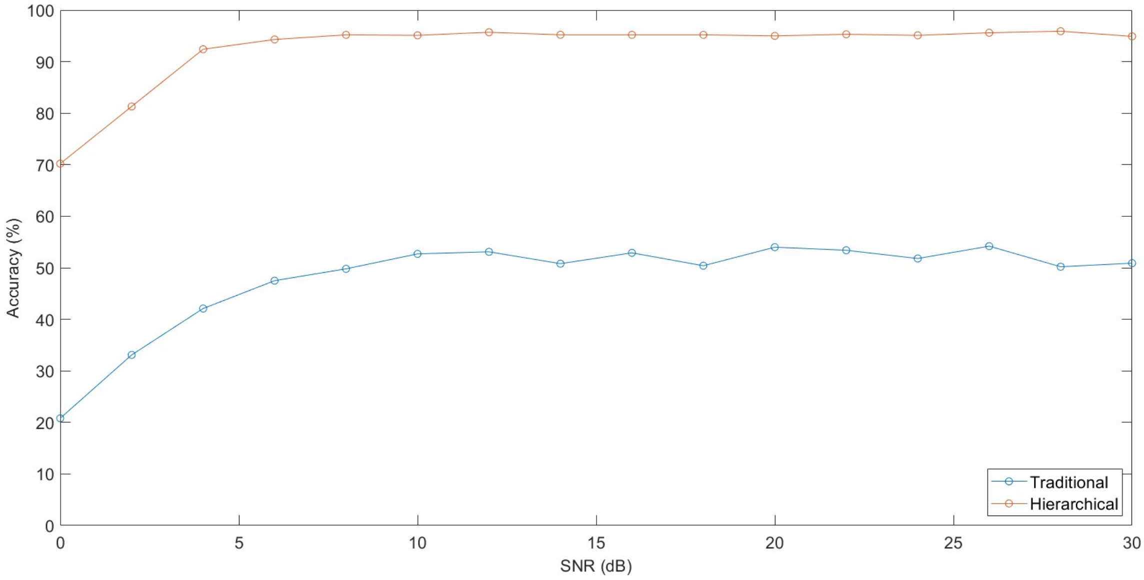

Figure 3.19: Accuracy for 128-APSK modulation

Since both traditional and hierarchical structures are insufficient to classify 8-PSK and 16-PSK signals, the overall accuracies are compared for remaining 10 modulation types as shown in Figure 3.20;

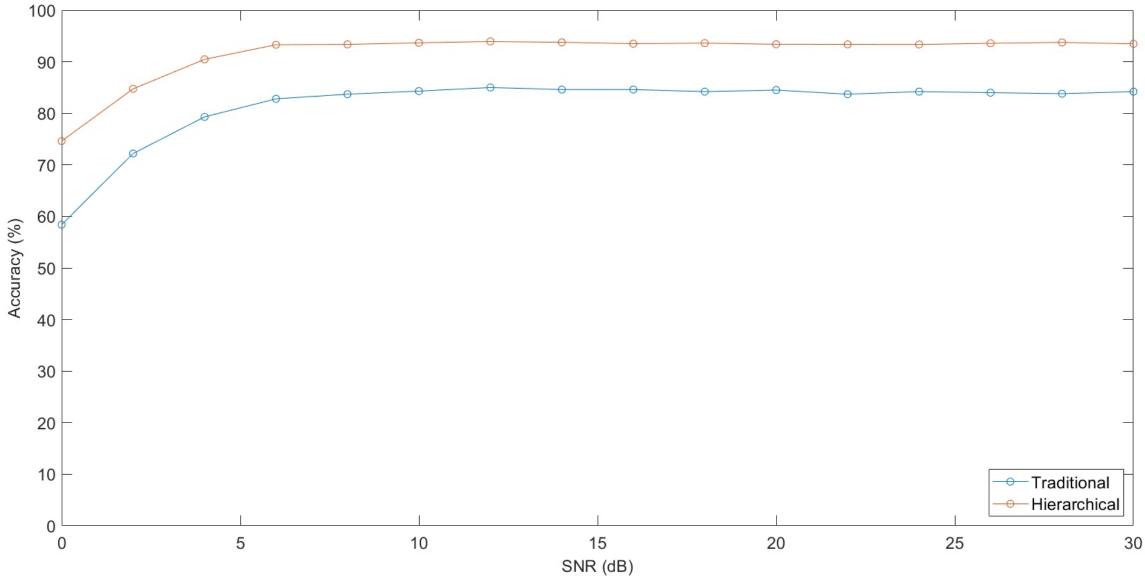

Figure 3.20: Overall accuracies for 10 modulations

CHAPTER 4

CONCLUSION

In this thesis, it is aimed to classify modulation types containing M-ASK, M-PSK, M-APSK up to higher orders at different SNR values using two different classification structures as traditional and hierarchical and to compare their performances. The classification process is carried out with various forms of SVM classifier associated with a feature-based approach. 41 different features are employed containing higher-order moments and cumulants up to 8^{th} order together with the mean, variance, skewness and kurtosis values of the signal characteristics.

Firstly, 12 different modulations given in Table 2.1 trained and tested on the classifiers linear, quadratic and cubic SVMs at several SNR values starting from 0 dB to 30 dB with 2 dB increasements. In this way, the SNR limits required for successful classification of these modulation types with using these features can be determined. The performances for these classifiers according to the SNR values are given in Figure 3.4. 80% of the dataset is used to train the classifier and 20% is used to test the trained classifier. As can be seen from the figure, from 0 dB until 6 dB linear and quadratic SVMs performed very close to each other. The performance of linear SVM is slightly lower than the others. After 8 dB all three classifiers showed the accuracy in the range of 70%-75%. Among them, quadratic SVM showed slightly higher performance. The confusion matrices for quadratic SVM at 0 dB, 10 dB and 20 dB are given in the Figures 3.5, 3.6 and 3.7 respectively. When the confusion matrices are examined, it can be seen that OOK and BPSK signals are not affected by the SNR variations. They show more than 90% success at all SNRs. Since they are binary modulations, it is much more harder that their constellations to confuse at low SNR values. At high SNR values, QPSK, 16APSK and 32APSK signals can be separated with high rates. In contrast, at low SNR values, they perform poorly and they are confused with other high order PSK and APSK signals. Since achieving high performances with high order modulations requires high SNR values, the SNR has an enormous effect on classifying these modulations. 8PSK, 16PSK and 32PSK signals are confused with each other as well as 64APSK and 128APSK at all SNR values. This structure is insufficient to distinguish these modulation types. It is recommended to apply different methods to classify them.

A hierarchical method is proposed to classify especially 4-ASK, 32-PSK, 16-APSK, 64-APSK, 128-APSK modulation types with higher performance. In the method using a single classifier trained with all 12 modulation types called traditional method, especially 32-PSK, 64-APSK and 128-APSK signals showed very poor performance. With using the proposed hierarchical method a significant increment in the performance of these modulation types is observed when comparing with the traditional method. When hierarchical methods are examined in the literature, classifiers are first classified according to modulation types and then reclassified according to modulation orders with several sub-classifiers and stages. This classification is carried out through numerous classification processes (at least five classifiers) [27], [37] which has a negative effect in terms of time and complexity or it has been applied to a limited number of modulation types [56], [57]. In this thesis, with only two stages, a total of three classifiers are used and a significant improvement in the classification rate is obtained. Since the classification rate increases in the hierarchical method, acceptable performance can also be obtained at SNR values less than 0 dB.

Among the all 12 modulations both traditional and hierarchical classification structures are failed to classify 8-PSK and 16-PSK modulations. Since both of these structures are insufficient to classify these two modulations, the classification performances of the remaining 10 modulation types are compared for these two structures in Figure 3.20. As shown in the figure, using the hierarchical structure brings approximately 15% increment in the overall accuracy. While the overall success rate of the traditional structure is around 80%, for hierarchical one it is approximately 95%.

As future work, for more comprehensive research, the dataset size and the variety of modulation types can be increased. For different modulation types, hierarchical classification structures that operate more efficiently at lower SNR levels can be examined. Also, classification using deep learning can be considered as a worthy future work of research.

REFERENCES

[1] H. Simon, *Communication systems*. Springer Nature, 1999.

[2] A. Goldsmith, *Wireless communications*. Cambridge university press, 2005.

[3] B. Sklar and F. J. Harris, *Digital communications: fundamentals and applications*, vol. 2001. Prentice-hall Englewood Cliffs, NJ, 1988.

[4] S. Haykin, *An introduction to analog and digital communication*. John Wiley, 1994.

[5] E. Azzouz and A. K. Nandi, *Automatic modulation recognition of communication signals*. Springer Science & Business Media, 2013.

[6] Z. Zhu and A. K. Nandi, *Automatic modulation classification: principles, algorithms and applications*. John Wiley & Sons, 2015.

[7] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 47–52, 2007.

[8] J. Hamkins, M. K. Simon, and J. H. Yuen, *Autonomous software-defined radio receivers for deep space applications*, vol. 13. Wiley Online Library, 2006.

[9] R. v. Nee and R. Prasad, *OFDM for wireless multimedia communications*. Artech House, Inc., 2000.

[10] V. Orlić and M. L. Dukić, "Algorithm for automatic modulation classification in multipath channel based on sixth-order cumulants," in *2009 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services*, pp. 423–426, IEEE, 2009.

[11] K. E. Nolan, L. Doyle, P. Mackenzie, and D. O'Mahony, "Modulation scheme classification for 4g software radio wireless networks," in *Proc. IASTED*, pp. 4–10, 2002.

[12] D. Grimaldi, S. Rapuano, and L. De Vito, "An automatic digital modulation classifier for measurement on telecommunication networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 5, pp. 1711–1720, 2007.

[13] K. E. Nolan, L. Doyle, D. O'Mahony, and P. Mackenzie, "Signal space based adaptive modulation for software radio," in *2002 IEEE Wireless Communications and Networking Conference Record. WCNC 2002 (Cat. No. 02TH8609)*, vol. 1, pp. 510–515, IEEE, 2002.

[14] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Survey of automatic modulation classification techniques: classical approaches and new trends," *IET communications*, vol. 1, no. 2, pp. 137–156, 2007.

[15] F. Hameed, O. A. Dobre, and D. C. Popescu, “On the likelihood-based approach to modulation classification,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5884–5892, 2009.

[16] J. L. Xu, W. Su, and M. Zhou, “Likelihood-ratio approaches to automatic modulation classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 455–469, 2010.

[17] W. Wei and J. M. Mendel, “Maximum-likelihood classification for digital amplitude-phase modulations,” *IEEE transactions on Communications*, vol. 48, no. 2, pp. 189–193, 2000.

[18] C.-Y. Huan and A. Polydoros, “Likelihood methods for mpsk modulation classification,” *IEEE Transactions on Communications*, vol. 43, no. 2/3/4, pp. 1493–1504, 1995.

[19] P. Panagiotou, A. Anastasopoulos, and A. Polydoros, “Likelihood ratio tests for modulation classification,” in *MILCOM 2000 Proceedings. 21st Century Military Communications. Architectures and Technologies for Information Superiority (Cat. No. 00CH37155)*, vol. 2, pp. 670–674, IEEE, 2000.

[20] O. Ozdemir, R. Li, and P. K. Varshney, “Hybrid maximum likelihood modulation classification using multiple radios,” *IEEE Communications Letters*, vol. 17, no. 10, pp. 1889–1892, 2013.

[21] F. Wang and X. Wang, “Fast and robust modulation classification via kolmogorov-smirnov test,” *IEEE Transactions on Communications*, vol. 58, no. 8, pp. 2324–2332, 2010.

[22] A. K. Nandi and E. E. Azzouz, “Algorithms for automatic modulation recognition of communication signals,” *IEEE Transactions on communications*, vol. 46, no. 4, pp. 431–436, 1998.

[23] H. Deng, M. Doroslovacki, H. Mustafa, J. Xu, and S. Koo, “Instantaneous feature based algorithm for hf digital modulation classification,” in *Proc. CISS Conf*, 2003.

[24] L. Hong and K. Ho, “Identification of digital modulation types using the wavelet transform,” in *MILCOM 1999. IEEE Military Communications. Conference Proceedings (Cat. No. 99CH36341)*, vol. 1, pp. 427–431, IEEE, 1999.

[25] C. Schreyogg, K. Kittel, U. Kressel, and J. Reichert, “Robust classification of modulation types using spectral features applied to hmm,” in *MILCOM 97 MILCOM 97 Proceedings*, vol. 3, pp. 1377–1381, IEEE, 1997.

[26] B. G. Mobasseri, “Constellation shape as a robust signature for digital modulation recognition,” in *MILCOM 1999. IEEE Military Communications. Conference Proceedings (Cat. No. 99CH36341)*, vol. 1, pp. 442–446, IEEE, 1999.

[27] A. Swami and B. M. Sadler, “Hierarchical digital modulation classification using cumulants,” *IEEE Transactions on communications*, vol. 48, no. 3, pp. 416–429, 2000.

[28] O. A. Dobre, Y. Bar-Ness, and W. Su, “Higher-order cyclic cumulants for high order modulation classification,” in *IEEE Military Communications Conference, 2003. MILCOM 2003.*, vol. 1, pp. 112–117, IEEE, 2003.

[29] K. Kim, I. A. Akbar, K. K. Bae, J.-S. Um, C. M. Spooner, and J. H. Reed, “Cyclostationary approaches to signal detection and classification in cognitive radio,” in *2007 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks*, pp. 212–215, IEEE, 2007.

[30] H. Alharbi, S. Mobien, S. Alshebeili, and F. Alturki, “Automatic modulation classification of digital modulations in presence of hf noise,” *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, p. 238, 2012.

[31] S. Xi and H.-C. Wu, “Robust automatic modulation classification using cumulant features in the presence of fading channels,” in *IEEE Wireless Communications and Networking Conference, 2006. WCNC 2006.*, vol. 4, pp. 2094–2099, IEEE, 2006.

[32] F. Liedtke, “Computer simulation of an automatic classification procedure for digitally modulated communication signals with unknown parameters,” *Signal processing*, vol. 6, no. 4, pp. 311–323, 1984.

[33] H. Gang, L. Jiandong, and L. Donghua, “Study of modulation recognition based on hocs and svm,” in *2004 IEEE 59th Vehicular Technology Conference. VTC 2004-Spring (IEEE Cat. No. 04CH37514)*, vol. 2, pp. 898–902, IEEE, 2004.

[34] J. Yuan, Z. Zhao-Yang, and Q. Pei-Liang, “Modulation classification of communication signals,” in *IEEE MILCOM 2004. Military Communications Conference, 2004.*, vol. 3, pp. 1470–1476, IEEE, 2004.

[35] M. W. Aslam, Z. Zhu, and A. K. Nandi, “Automatic modulation classification using combination of genetic programming and knn,” *IEEE Transactions on wireless communications*, vol. 11, no. 8, pp. 2742–2750, 2012.

[36] E. E. Azzouz and A. K. Nandi, “Modulation recognition using artificial neural networks,” in *Automatic Modulation Recognition of Communication Signals*, pp. 132–176, Springer, 1996.

[37] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, “Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers,” *Physical Communication*, vol. 21, pp. 10–18, 2016.

[38] G. Hatzichristos and M. P. Fargues, “A hierarchical approach to the classification of digital modulation types in multipath environments,” in *Conference Record of Thirty-Fifth Asilomar Conference on Signals, Systems and Computers (Cat. No. 01CH37256)*, vol. 2, pp. 1494–1498, IEEE, 2001.

[39] B. Ramkumar, *Automatic Modulation Classication and Blind Equalization for Cognitive Radios*. PhD thesis, Virginia Tech, 2011.

[40] E. E. Azzouz and A. K. Nandi, “Automatic identification of digital modulation types,” *Signal Processing*, vol. 47, no. 1, pp. 55–69, 1995.

[41] A. Swami and B. Sadler, "Modulation classification via hierarchical agglomerative cluster analysis," in *First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, pp. 141–144, IEEE, 1997.

[42] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[43] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[45] T. N. Sainath, R. J. Weiss, A. Senior, K. W. Wilson, and O. Vinyals, "Learning the speech front-end with raw waveform cldnns," in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[46] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, 2018.

[47] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *International conference on engineering applications of neural networks*, pp. 213–226, Springer, 2016.

[48] R. W. Klein, M. A. Temple, and M. J. Mendenhall, "Application of wavelet-based rf fingerprinting to enhance wireless network security," *Journal of Communications and Networks*, vol. 11, no. 6, pp. 544–555, 2009.

[49] "What is on-off keying modulation?." `https://www.quora.com/What-is-on-off-keying-modulation?redirected_qid=55386904`. Accessed: 24.12.2019.

[50] L. Liu and J. Xu, "A novel modulation classification method based on high order cumulants," in *2006 International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–5, IEEE, 2006.

[51] A. Nandi and E. E. Azzouz, "Automatic analogue modulation recognition," *Signal processing*, vol. 46, no. 2, pp. 211–222, 1995.

[52] N. P. Geisinger, "Classification of digital modulation schemes using linear and nonlinear classifiers," tech. rep., NAVAL POSTGRADUATE SCHOOL MONTEREY CA, 2010.

[53] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[54] "Deepsig." `https://www.deepsig.io/datasets`. Accessed: 12.12.2019.

[55] T. W. Rondeau, *Application of artificial intelligence to wireless communications*. PhD thesis, Virginia Tech, 2007.

[56] F. Xiang and L. Jiandong, "A hierarchical digital modulation classification algorithm for adaptive wireless communication systems," *Wireless Personal Communications*, vol. 39, no. 3, pp. 321–326, 2006.

[57] S. Majhi, R. Gupta, W. Xiang, and S. Glisic, "Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11057–11069, 2017.