LEARNING-BASED AUTOMATIC MODULATION CLASSIFICATION

by

Ameen Elsiddig Abdelmutalab

A Thesis Presented to the Faculty of the
American University of Sharjah
College of Engineering
in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science in
Electrical Engineering

Sharjah, United Arab Emirates

May 2015

Approval Signatures

We, the undersigned, approve the Master's Thesis of Ameen Elsiddig Abdelmutalab.

Thesis Title: Learning-Based Automatic Modulation Classification.

Signature	**Date of Signature** (dd/mm/yyyy)
_____	_____
Dr. Khaled Assaleh Professor, Department of Electrical Engineering Thesis Advisor	
_____	_____
Dr. Mohamed El-Tarhuni Professor, Department of Electrical Engineering Thesis Advisor	
_____	_____
Dr. Arafat Al-Dweik Associate Professor, Department of Electrical Engineering Khalifa University Thesis Committee Member	
_____	_____
Dr. Usman Tariq Associate Professor, Department of Electrical Engineering American University of Sharjah Thesis Committee Member	
_____	_____
Dr. Nasser Qaddoumi Head, Department of Electrical Engineering	
_____	_____
Dr. Mohamed El-Tarhuni CEN Graduate Programs Director	
_____	_____
Dr. Leland Blank Dean, College of Engineering	
_____	_____
Dr. Khaled Assaleh Interim Vice Provost for Research and Graduate Studies	

Acknowledgements

I would like to express my sincere gratitude to my advisors, Dr. Khaled Assaleh and Dr. Mohamed El-Tarhuni, for their guidance, motivation and endless support. I would also like to thank the American University of Sharjah for this opportunity to be part of their graduate program and granting me a graduate teaching assistantship.

My sincere thanks go to my parents for their continuous support, not to forget their prayers and kind words that helped me throughout my way. I am also grateful to my friends Hesham Gamal, Adil Khurram, Menatalla Shehab, and Sana Hamid. Last but not least, I would also like to express my sincere thanks to all my friends with whom I shared my best memories.

To those who lost their

Right to choose . . .

Abstract

Automatic Modulation Classification (AMC) is a new technology implemented into communication receivers to automatically determine the modulation type of a received signal. One of the main applications of AMC is in adaptive modulation systems, where the modulation scheme is changed dynamically according to the changes in the wireless channel. However, this requires the receiver to be continuously informed about the modulation type, resulting in a loss of bandwidth efficiency. The existence of smart receivers that can automatically recognize the modulation type improves the utilization of available bandwidth. In this thesis, a new AMC algorithm based on a Hierarchical Polynomial Classifier structure is introduced. The proposed system is tested for classifying BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM modulation types in Additive White Gaussian Noise (AWGN) and flat fading environments. Moreover, the system uses High Order Cumulants (HOCs) of the received signal as discriminant features to distinguish between the different digital modulation types. The proposed system divides the overall modulation classification problem into hierarchical binary sub-classification tasks. In each binary sub-classification, the HOC inputs are expanded into a higher dimensional space in which the two classes are linearly separable. Furthermore, the signal-to-noise ratio of the received signal is estimated and fed to the proposed classifier to improve the classification accuracy. Another modification is added to the proposed system by using stepwise regression optimization for feature selection. Hence, the input features to the classifier are chosen to give the highest classification accuracy while maintaining a minimum number of possible features. Extensive simulations showed that a significant improvement in classification accuracy and reduction in the system complexity is obtained compared to the previously suggested systems in the literature.

Search Terms: ***Adaptive Modulation, Automatic Modulation Classification (AMC), Hierarchical Polynomial Classifiers (HPC), SNR Estimation, Stepwise Regression.***

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALRT Average Likelihood Ratio Tests.

AMC Automatic Modulation Classification.

ANN Artificial Neural Networks.

ASK Amplitude Shift Keying.

AWGN Additive White Gaussian Noise.

BAT Bit Allocation Table.

BER Bit Error Rate.

BPSK Binary Phase Shift Keying.

CR Cognitive Radio.

FSK Frequency Shift Keying.

GLRT Generalized Likelihood Ratio Tests.

HLRT Hybrid Likelihood Ratio Tests.

HOCs High Order Cumulants.

HOMs High Order Moments.

HPC Hierarchical Polynomial Classifier.

KNN K-Nearest Neighbor.

LOS Line of Sight.

OFDM Orthogonal Frequency Division Multiplexing.

PDF Probability Density Function.

PSD Power Spectral Density.

QAM Quadrature Amplitude Modulation.

QoS Quality of Service.

QPSK Quadrature Phase Shift Keying.

SNR Signal-to-Noise Ratio.

SVM Support Vector Machines.

Chapter 1: Introduction

Automatic Modulation Classification (AMC) has been a topic of research for the last few decades due to its many applications in modern communication systems. For example, in adaptive modulation, the modulation type of the transmitted signal is changed based on the state of the channel between the transmitter and receiver, where high modulation levels are used whenever the channel is clean, and low modulation levels are used for noisy channels. The former procedure enhances the use of the available spectrum while maintaining a low probability of error. However, any modulated signal needs to be demodulated at the receiver side in order to extract the original transmitted messages. This is usually performed by sending a pilot signal along with the original message to inform the receiver about the modulation type of the transmitted signal. Although the previous solution is valid and can solve the problem, it has a number of drawbacks. Firstly, it reduces the data throughput of the system by periodically sending pilot signals that does not carry data, especially for systems with very small data blocks. Moreover, in many practical life applications, such as military applications, there is no prior agreement between the transmitter and receiver. Automatic modulation classification solves this problem by using some properties of the received signal in order to identify the modulation identity. The challenge is to select the proper discriminating properties (features) that can separate the different modulation types. In this thesis, we develop a new automatic modulation classification system that achieves high classification accuracy while maintaining a simple structure, which is an important feature for many practical applications. Most of the proposed systems in the literature assume perfect information about the channel at the receiver side, which is not a valid assumption all the time. In this work, a new SNR estimation algorithm is proposed, where the estimated SNR value is fed to the modulation classifier in order to improve its overall classification accuracy. Moreover, modulation classification is usually a real-time problem that should be solved within a specified, relatively short, time constraint. Therefore, in order to minimize the complexity of the proposed system, a feature selection algorithm is proposed resulting in a significant reduction in the order of required calculations.

1.1. Contribution

The contribution of this thesis is summarized in the following points:

- Propose an automatic modulation classification system based on Hierarchical Polynomial Classifier (HPC) and High Order Cumulants (HOCs). The proposed system has the following advantages:
  1. Provides higher classification accuracy compared to most of the proposed systems in the literature.
  2. Has low computational complexity and can be easily implemented.
  3. Is robust to rotations in the signal constellation.
  4. Able to identify different modulation schemes even in the existence of Rician or Rayleigh fading.
- An optimization for the number of received symbols used in extracting the classification features is carried out in order to provide the best classification accuracy for flat fading channels.
- Feature selection system based on stepwise regression is integrated with the proposed Hierarchical Polynomial Classifier resulting in a simplified final classifier model.
- An SNR estimation technique is proposed, the system is able to recognize the SNR of the signal with high accuracy, especially at low SNR values.
- Different classification algorithms are simulated and examined, and their performance and complexity are investigated.

1.2. Thesis Outline

This thesis is organized as follows: **Chapter 1** presents the motivation for research, and thesis contribution. **Chapter 2** delivers an important background in order to fully understand the problem and its circumstances, including the building blocks of communication systems, modulation schemes and models of wireless communications channels. **Chapter 3** discusses the previous work done in the literature. The chapter starts by highlighting the known applications of AMC. Then a brief explanation of the

different proposed features in the literature is presented. Finally, some optimization algorithms are introduced such as Genetic Programing and Particle Swarms Optimization. **Chapter 4** introduces an overview to pattern recognition systems and presents the structure and mechanism of different machine learning classifiers. **Chapter 5** presents the proposed model including the polynomial classifier, hierarchical polynomial classifier, SNR estimation model, and stepwise regression system. **Chapter 6** includes the simulation results and provides a comparison between the proposed scheme and other related work. **Chapter 7** concludes the thesis and discusses the thesis outcomes and the future work.

Chapter 2: Background

In this chapter, the end-to-end components of communication systems are introduced, and a brief review of M-ary Phase Shift Keying (MPSK) and M-ary Quadrature Amplitude Modulation (MQAM) and their constellations is presented. Moreover, different types of signal impairments including noise and fading are covered. Finally, the effect of different Doppler shift values on the received constellation is discussed.

2.1. Digital Communication Systems

The use of digital communication systems is rapidly growing due to the increasing need for fast and reliable data transmission. However, the performance of such systems affects the quality of the data received by the end user. For example, real-time applications such as video streaming and voice over IP are very sensitive to delays in transmission. Accordingly, increasing the data rate is necessity in order to meet the minimum quality of service (QoS) expected by the user. Furthermore, technological advancements in various fields are becoming more dependent on wireless communication systems given the availability of high transmission rates with low error probability. All these reasons, in addition to the continuous growth of the number of users, motivated research that investigates new solutions and optimizes the available bandwidth resources to satisfy all the growing needs [1].

Figure 1 shows the basic components of a typical communication system. First, input messages such as voice, text, and video are converted by an input transducer in order to form the electrical input signal to the transmitter. Then, the signal is modulated and conveyed through wired or wireless communication channel. At the receiver side, the demodulation process is performed and the transmitted signal is extracted; then the output transducer converts the received signal into its original form [2].

2.2. Modulation

The baseband signals at the transmitter are usually low frequency signals with very large wavelengths. Emitting this type of signals requires a radiating antenna with

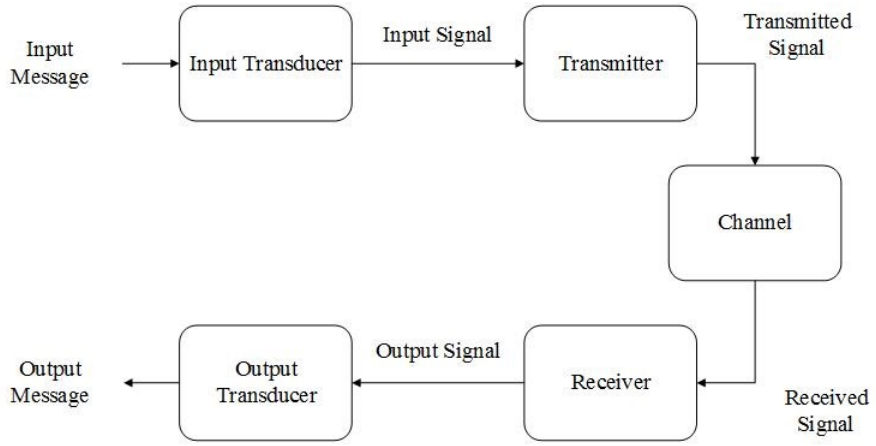


Figure 1: Block digram of a typical communication system.

enormous dimensions (antenna dimensions are related to the wavelength of the signal) [2]. Furthermore, most of the original low frequency signals occupy almost the same bandwidth frequency; therefore transmitting them as they are results in overlapping between the signals and loss of information.

The modulation process modulates the baseband signal information onto a higher frequency signal called the carrier. This process transfers the frequency of the signal being transmitted to a new center frequency that occupies a reserved bandwidth in the spectrum. Thus, the transmission of the signal occurs without any interference with other signals (in the optimal case scenario). Modulation can be performed by changing different parameters of the carrier signal based on the nature of the original baseband signal being transmitted. For example, Amplitude Shift Keying (ASK) modulation changes the amplitude of the carrier signal in order to convey the transmitted message, where each amplitude level represents a different transmitted sample. As for Frequency Shift Keying (FSK) modulation, each symbol of the signal is coded as a certain frequency; hence, the transmitted signal contains more than one frequency. Another type of modulation that is widely used modifies the phase of the carrier signal based on the signal being modulated into it. This type of modulation is called Phase Shift Keying (PSK). Finally, a hybrid modulation techniques called Quadrature Amplitude Modulation (QAM), which is a mix between PSK and ASK modulation, carries the transmitted symbols in both the amplitude and the phase of the carrier. However, in this work we will consider different levels of PSK and QAM modulations.

2.2.1. Phase Shift Keying.

In Phase Shift Keying (PSK), the baseband symbol is conveyed by the phase of the carrier signal. For example, Binary Phase-Shift Keying (BPSK) uses two different phases to represent the value of the transmitted symbol (either 0 or 1), while Quadrature Phase-Shift Keying (QPSK) uses four different phases to represent the two transmitted symbols of the signal (00, 01, 10 and 11).

In general, higher-order PSK modulations transmit more bits per symbol; thus they achieve higher data rate transmission, a very attractive feature for all commutation systems. Unfortunately, although PSK modulation comes with multiple benefits, the associated probability of error increases as the level of modulation increases, i.e. 16-PSK has a higher probability of error compared to the simple BPSK. However, the main factor in favoring one modulation level over the other is the quality of service (QoS) required for each application. Some applications, such as video streaming and voice calls, are very sensitive to time delay but are less sensitive to errors in the transmission; therefore, higher modulation levels are recommended. On the other hand, applications like emails and banking services are more sensitive to transmission errors; for such applications, lower modulation schemes can be used to minimize the probability of errors.

Higher-order PSK modulations are usually referred to as M-ary PSK, The wave form of which is expressed as:

$$S_i(t) = A\cos(\omega_c t + \phi_i) \qquad i = 0,1,2,...,M-1 \qquad 0 \leqslant t \leqslant T \tag{1}$$

where

$$A = \sqrt{\frac{2E}{T_s}}, \;\; \phi_i = \frac{2\pi i}{M} \qquad i = 0,1,2,...,M-1 \tag{2}$$

and where E is the symbol energy, T_s is the symbol time, $0 \leq t \leq T_s$, ϕ_c is the carrier frequency and M is the number of different possible symbols in the M-ary PSK modulation scheme. For example, in BPSK modulation, M=2 with only two possible transmitted symbols and bandwidth efficiency of a 1 $bit/s/Hz$, and for QPSK modulation, M=4 with a bandwidth efficiency of 2 $bit/s/Hz$. Figure 2 shows the 8-PSK constellation with M=8, and each symbol represents 3 different transmitted bits.

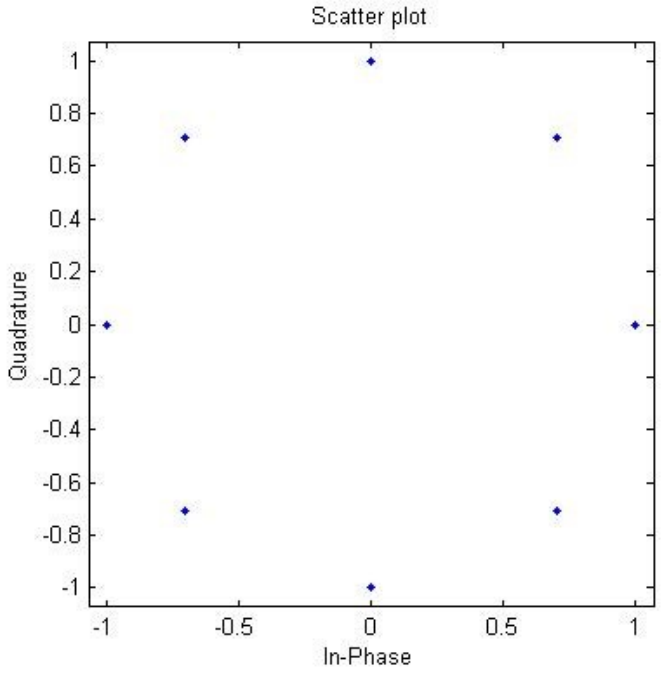


Figure 2: 8-PSK constellation.

2.2.2. Quadrature Amplitude Modulation.

Quadrature Amplitude Modulation (QAM) is a modulation scheme that conveys the baseband signal in both amplitude and phase of the carrier. It is a combination of ASK and PSK modulations. Unlike PSK, QAM symbols are represented by different levels of energy. M-ary QAM waveform can be represented as:

$$S_i = A_{mi} cos(\omega_c t) - A_{mq} sin(\omega_c t) \qquad i = 0, 1, 2, ..., M-1 \tag{3}$$

where A_{mi} and A_{mq} are the signal amplitude in the in-phase and the quadrature components, respectively, and M is the number of different possible symbols in the M-ary QAM modulation scheme. Generally, QAM modulation has different forms; however, in this work the discussion is narrowed to only include square-shaped QAM.

High-order QAM allows for high data rates and improved bandwidth efficiency. However, it is easily corrupted by noise and interference. Hence, the modulation type has to be optimized to yield the highest possible data rate and still abide by the limiting BER specified by the system; i.e. the modulation type can be chosen to make the best use of the available resources [3].

Figure 3 shows the constellation diagram for a 64-QAM with 64 different possible symbols; each symbol represents 6 different bits. In general, for any MPSK and MQAM, the number of bits per symbol is given by:

$$k = \log_2(M) \tag{4}$$

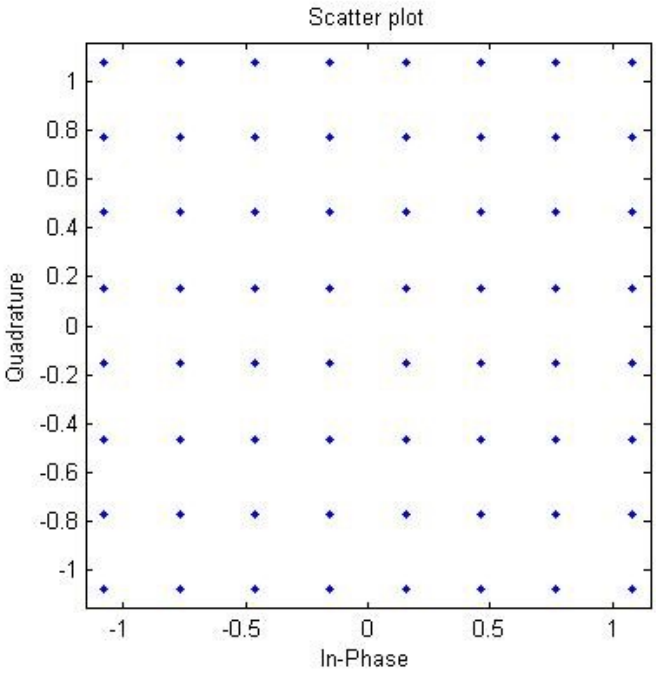


Figure 3: 64-QAM constellation.

2.3. Noise

Noise is one of the dominant impairments in wireless communication systems. Noise can be due to the thermal energy in the receiver emitted by its antenna or an other component of the receiver circuit, or because of the nonlinear properties of some filters and devices (intermodulation noise). Noise corrupts the received signal and results in errors when extracting the original message.

Figure 4 shows the Power Spectrum Density (PSD) of Additive White Gaussian

Noise (AWGN) channels; the power density has constant magnitude in all the frequencies. The effect of noise on the transmitted signal can be described by a ratio between signal power and noise power (SNR), or sometimes by a ratio between the symbol energy E_s and noise power spectrum density N_0:

$$SNR = \frac{E_s}{N_0} \tag{5}$$

In practical applications, noise is not the only impairment that undergoes the power of transmitted signals and causes errors. A severe impact on the transmitted signal is caused by the fading phenomena, where multiple copies of the same signal arrive at the receiver side. However, scientists introduced different models in order to study the performance of communication systems. The simplest model is the free space propagation model which is introduced in the next section.

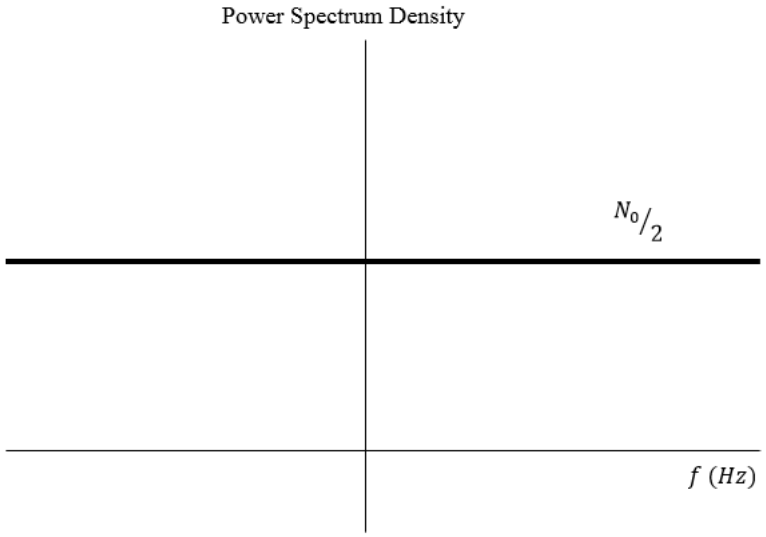


Figure 4: Power spectral density for white noise.

2.4. Free Space Propagation Model

This model is used to describe the power of the received signal in free space environments with no reflections between the transmitter and the receiver. The received power is inversely proportional to the square of the distance between the transmitter and

the receiver:

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} \tag{6}$$

where P_r is the received power, P_t is the transmitted power, G_t and G_r are the transmitter and the receiver antenna gains, respectively, λ is the wavelength of the transmitted signal, and d is the distance between the transmitter and the receiver.

2.5. Fading

To study the performance of any communication system, the wireless channels are usually assumed to be ideal where the transmitted signals are only corrupted by AWGN noise. This assumption gives a basic understanding of the behavior of systems and reflects their major trends. However, the interference between different signals in the spectrum or between different copies of the same signal has a major impact and can cause a significant degradation in the performance of the system [4]. In general, the channel between the transmitter and the receiver varies with time and frequency. This variation can be divided into:

- **Large-Scale Fading:** This type of fading describes the degradation in the signal power with the distance or due to the shadowing effect by buildings or mountains. Large-Scale Fading takes place in large distances where all the signals with different frequencies experience the same behavior (are frequency independent).
- **Small-Scale Fading:** This type of fading happens in small distances, usually for distances in the order of the signal wavelength. It is normally due to multipath phenomena where reflected copies of the same signal arrive at the receiver and add together constructively or destructively [5]. Figure 5 shows an example of multipath fading.

2.5.1. Small-Scale Fading.

Based on the multipath time delay spread of the channel, small scale fading can be divided into two types, either flat fading or frequency selective fading:

Flat Fading.

Flat fading is the type of fading in which the signal is transmitted over a band-

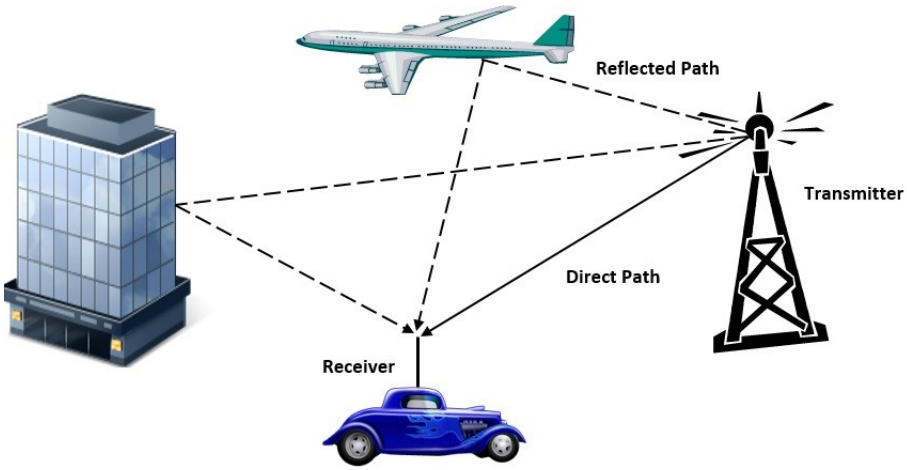


Figure 5: Multipath fading.

width with a constant gain, which means all the frequency parts of the transmitted signal experience the same fading and are multiplied by the same gain. Flat fading takes place when:

$$B_s \ll B_c \quad and \quad T_s \gg \sigma_\tau \tag{7}$$

where B_s is the bandwidth of the transmitted signal and B_c is the coherence bandwidth of the channel. This can also be expressed in terms of the symbol time of the transmitted signal T_s, and the channel spread delay σ_τ, meaning that all the reflections of the transmitted message arrive at the receiver within the symbol period of the signal, resulting in no interference between the consecutive transmitted symbols [6].

Frequency Selective Fading.

Frequency selective fading has a more severe impact on the system performance compared to flat fading. For frequency selective fading, the signal bandwidth is greater than the coherence bandwidth of the channel. In other words, the symbol interval of the signal is smaller than the delay spread of the channel. Thus, the reflected copies of the same signal arrive at the receiver during the time of receiving the next symbol or maybe the symbols after that. This type of fading causes severe inter-symbol interference (ISI) between the consecutive received symbols. An equalizer is required at the receiver side to compensate for the changes produced by the fading nature of the

channel. Mathematically, frequency selective fading takes place when:

$$B_s \gg B_c \quad and \quad T_s \ll \sigma_\tau \tag{8}$$

Furthermore, small scale fading can also be divided, based on the Doppler Spread of the channel, into fast or slow fading:

Fast Fading.

This type of fading occurs when the relative motion between the transmitter and the receiver is high. Accordingly, the channel impulse response is rapidly changing during the symbol period. Thus, the transmitted symbols experience different channel gains which distort the final received signal. The coherence time of a channel T_D is defined as the time over which the channel impulse response is constant. If the coherence time of the channel is smaller than the symbol period, then the signal is undergoing fast fading [7]. This can also be written in terms of the bandwidth of the transmitted signal B_s and the *Doppler spread*B_D as:

$$B_s \ll B_D \quad and \quad T_s \gg T_D \tag{9}$$

Slow Fading.

This type of fading has less impact on the performance of wireless communication systems, since the channel impulse response is constant during the time of transmitting the symbol. Therefore, the transmitted symbol is multiplied by a constant gain without causing any intersymbol interference between the adjacent transmitted symbols. The condition for slow fading is:

$$B_s \gg B_D \quad and \quad T_s \ll T_D \tag{10}$$

2.5.2. Rayleigh Fading.

The Rayleigh fading model represents the random change in the magnitude of the received signal due to the arrival of multiple copies of the same signal at the receiver

when there is no clear path between the transmitter and the receiver. The complex phasor of the received signal envelope due to N copies of the reflected signal is expressed as:

$$\tilde{E} = \sum_{n=1}^{N} E_n e^{j\Theta_n} \tag{11}$$

where E_n and $e^{j\Theta_n}$ are the relative magnitude and phase of each path, respectively. The values of Θ_n are assumed to be independent and uniformly distributed, since the phase magnitude is very sensitive to small changes in the path length. According to the Central Limit Theorem, the summation in (11) can be represented by a Gaussian distribution for a large number of independent reflected paths. Accordingly, the Probability Density Function (PDF) of the received signal envelope is expressed as [8].

$$P(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \qquad r \geqslant 0 \tag{12}$$

where σ is the rms value of the reflected signals. Rayleigh Fading is the most used model for wireless channels in urban environments due to the blocking of radio signals by buildings.

2.5.3. Rician Fading.

In the Rayleigh fading model, reflected signals from different paths are assumed to have relatively the same power. In other words, there is no Line of Sight (LOS) path between the transmitter and the receiver. For some scenarios, like indoor wireless communication, a LOS path between transmitter and the receiver exists; hence, the Rayleigh fading model is no longer accurate to describe the distribution of the amplitude of the received signal. The Rician fading channel model assumes that reflected signals along with a direct signal arrive at the receiver. Therefore, the complex phasor of the received signal envelope is expressed as:

$$\tilde{E} = E_0 + \sum_{n=1}^{N} E_n e^{j\Theta_n} \tag{13}$$

where E_0 is the LOS component of signal. Moreover, Rician PDF is modeled as:

$$P(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+v^2}{2\sigma^2}} I_0(\frac{rv}{\sigma^2}), \qquad r \geqslant 0 \tag{14}$$

where v^2 is the power in the non-fading component and I_0 is the modified Bessel function of zeroth order. Another important factor when defining the Rician fading is the power of the direct path with respect to the reflected paths which can be determined by the K factor as:

$$K = \frac{v^2}{\sum_{n=1}^{N} |E_n|^2} \tag{15}$$

where $|E_n|^2$ is the power of the n^{th} reflected path. In general, when the K factor increases, the probability of the signal to undergo a deep fading decreases.

2.6. Doppler Shift

Doppler shift is the change in the frequency of the received signal when either or both the transmitter or the receiver is in motion.The detected frequency at the receiver will be different than the original transmitted frequency based on the direction of movement. For instance, when the receiver is moving toward the transmitter, the received frequency appears higher than the original transmitted frequency. Likewise, if the receiver is moving away from the transmitter, the received frequency appears to be lower than the original transmitted frequency.

For example, when the receiver is moving with respect to the transmitter, the received frequency is:

$$f_r = f_0(1 \pm \frac{v}{c}) \tag{16}$$

where f_0 is the original transmitted frequency, v is the speed of the receiver, and c is the speed of light. Accordingly, the resultant Doppler shift can be expressed by:

$$f_D = f_r - f_0 = \pm f_0 \frac{v}{c} \tag{17}$$

Generally, if there is an angle Θ between the motion of the receiver and the

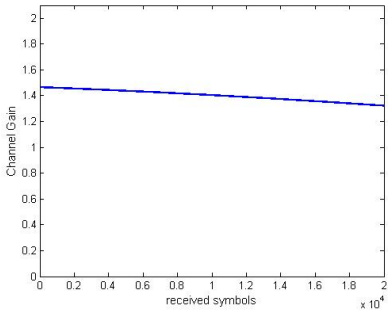
direction of radiation, the resultant Doppler shift is:

$$f_D = \pm f_0 \frac{v}{c} cos\Theta \tag{18}$$

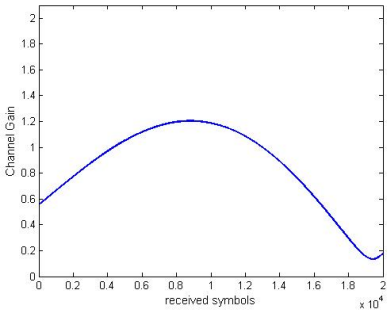
However, the maximum Doppler shift occurs when the angle between the transmitter and the receiver is zero. Thus, maximum Doppler shift can be written as:

$$f_{D(Max)} = \pm f_0 \frac{v}{c} \tag{19}$$

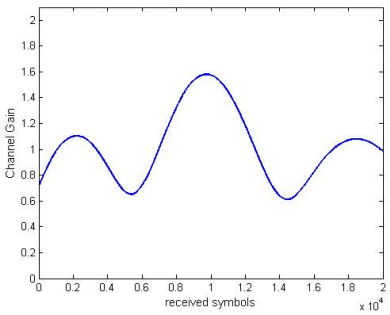
Figure 6 shows examples of channels with different values of Doppler shifts. The channels have two reflected paths (Rayleigh Fading) with Doppler shifts equal to 3Hz, 30Hz, and 83Hz, respectively. It is clear that, for higher Doppler shifts, the channel expresses more fluctuation, which means the received symbols experience channels with different gains. Accordingly, the received constellation appears very different compared to the original constellation of normal MPSK and MQAM.



(a) 3Hz Doppler Shift



(b) 30Hz Doppler Shift



(c) 83Hz Doppler Shift

Figure 6: Rayleigh channel with different values of Doppler shifts.

Chapter 3: Literature Review

In this chapter, we will introduce some applications of automatic modulation classification in cognitive radio, OFDM, and electronic welfare systems. We will also discuss various proposed algorithms for AMC including likelihood and pattern recognition based approaches. Moreover, different classification features are used in the literature such as Instantaneous Amplitude, Phase and Frequency, Spectrum Analysis, Wavelet Transform, Clustering Algorithms, and High Order Cumulants. We will go through these features in detail in order to show their calculations, advantages, disadvantages, and limitations. Finally, we will dedicate the last section of this chapter to introduce some feature optimization techniques like Genetic Programming and Particle Swarm Optimization.

3.1. Automatic Modulation Classification (AMC)

Automatic Modulation Classification (AMC) is the process of identifying the modulation type of a transmitted signal from the received data samples automatically [9]; it is an intermediate step between signal detection and demodulation. AMC has received a great deal of research and investigation in recent years because of its various applications in modern communication systems.

3.1.1. Applications.

Automatic modulation classification has many applications in cognitive radio, OFDM, and electronic welfare systems.

Cognitive Radio (CR).

In many communication systems, a reserved spectrum bandwidth is specified for the different licensed services. Studies showed that, most of the time the reserved bandwidth is unoccupied and significant available resources are wasted [10]. Cognitive Radio (CR) systems provide a solution for this problem by giving the priority of using the reserved spectrum to the primary users (licensed services), and allowing the secondary users (other users in the system) to use the spectrum if it is unoccupied. This

process optimizes the usage of the available spectrum, and allows both the primary and the secondary users to use the spectrum efficiently. However, the secondary users are assumed to be able to recognize the state of the spectrum (whether it is occupied or not). AMC is used to identify the types of signals in the spectrum and therefore would be implemented in the secondary transmitter [11].

Orthogonal Frequency Division Multiplexing (OFDM).

In wireless communication, channels are time varying and many obstacles exist between the transmitter and the receiver. Using a single carrier transmission method in such multipath fading environments results in a corrupted received signal, as the different frequencies in the transmitted signal experience different channel gains.

OFDM schemes solves this problem by dividing the spectrum into smaller sub-bands, then using a single sub-carrier for each sub-band. Consequently, all the small frequency bands are transmitted through flat fading channels and the effect of the inter-symbol interference is reduced.

Moreover, different modulation levels can be used based on the channel state of each sub-band [12]. This process is known as adaptive modulation. For example the standards OFDM protocol IEEE 802.11a has throughput in the range of 6 Mbps for BPSK modulation and 48 Mbps for 64-QAM modulation, but the probability of error increases when the level of modulation increases. Accordingly, high modulation levels are used for sub-carriers with high SNRs, and low modulation levels are used for sub-carriers with low SNRs, resulting in a significant improvement in the throughput of communication systems.

Receivers in adaptive modulation systems are required to identify the modulation type of each sub-carrier in order to select the corresponding demodulation method. This can be done using a Bit Allocation Table (BAT), which is sent within each transmitted frame in order to inform the receiver about the modulation type of each sub-carrier [13]. Unfortunately, the BAT adds an extra overhead, especially for high numbers of sub-carriers and small OFDM frames. An attractive solution is to use AMC at the receiver side to recognize the modulation type of each sub-carrier, therefore increasing the overall transmission rate of the system.

Other Applications.

Automatic modulation classification has many applications in civilian areas like spectrum surveillance, interference identification, transmission control and monitoring, and software defined radio (SDR). In addition, it has applications in military areas like: electronic warfare, threat analysis, target acquisition, jamming, and homing [14]. For example, in military applications, automatic modulation classification is used to identify the modulation type of an intercepted enemy's signal where there is no information available about the modulation type of such signals [15].

Different solutions were proposed to accomplish successful modulation classification. Generally we can divide modulation classification techniques into two main approaches:

- Maximum likelihood (ML).
- Feature-based methods.

3.2. Maximum Likelihood (ML)

Maximum likelihood classifiers formulate the modulation recognition process as multiple hypothesis testing problems, in which each modulation scheme is represented by one hypothesis. ML classifiers optimize the classification procedure and give the best possible recognition rate. This is performed by finding the likelihood function, which is usually a function of the transmitted symbols and channel parameters [16]. The candidate with the highest likelihood is chosen as the modulation type.

$$\hat{s} = \arg\max_{\mathbf{s}_m}(p(\mathbf{r}|\mathbf{s}_m)) \tag{20}$$

where $\hat{s}$ represents the selected modulation scheme by the receiver, $p(\mathbf{r}|\mathbf{s}_m)$ is the likelihood that the received signal is $\mathbf{r}$ given that the transmitted signal is $\mathbf{s}_m$, where $\mathbf{r}_m$ is modulated using modulation type m.

Maximum likelihood approaches can be divided into three main categories: Average Likelihood Ratio Tests (ALRT), Generalized Likelihood Ratio Tests (GLRT) and Hybrid Likelihood Ratio Tests (HLRT). In ALRT, the unknown signal quantities like signal constellation and noise power are treated as random variables where the proba-

bility density function (PDF) of the received signal is computed by averaging over these random variables with the assumption that their distributions are known. This method leads to the best classification accuracy if this assumption is accurate. In contrast, for GLRT, the PDF of the received signal is calculated using the maximum likelihood estimation (MLE) of the unknown values, where they are treated as deterministic variables. However, ALRT requires heavy calculations, especially when the number of random variables increases; on the other hand, GLRT is less complex but has a drawback when it is used to classify nested signal constellations like (16-QAM and 64-QAM) resulting in an incorrect classification. This problem is solved by HLRT, which has the benefits of both of ALRT and GLRT. Hence, HLRT assumes that some parameters of the signal are random variables with known PDFs and some of the parameters are deterministic [17]. Although ML approaches can lead to the optimal solution, feature extraction approaches are usually preferred due to their low complexity and satisfactory performance.

3.3. Feature-Based Method

In Feature-Based (FB) approaches, the modulation class is identified using a two-step process. First, representative features are used to represent the received signal instead of dealing with the signal as a stream of symbols. Then, the selected features are used by a machine learning classifier in order to make a decision about the modulation class. Feature extraction methods are simpler to implement and can lead to suboptimal solutions compared to the likelihood-based approach. The efficiency of this approach depends on the classification power of the selected features and the simplicity of the applied classifiers. Many types of features are proposed in the literature. However, we will dedicate this section to discussing these features and explaining their power and usage.

3.3.1. Spectrum Analysis-Based Classification Scheme.

Analyzing the spectrum of an unknown signal is used to identify certain modulation types. For example, Multiple Frequency Shift Keying (MFSK) signals convey baseband information into different carrier frequencies; the magnitude of the spectrum at these frequencies is higher than at other unused frequencies. Likewise, the phase

spectrum of MPSK determines the number of phases used to modulate the baseband signal. Generally, frequency and phase spectrums cannot distinguish all types of modulation, and their performance varies from one group of modulation to another [18].

In [19], the Fast Fourier Transform (FFT) of the received signal is used to identify its modulation type. For example, if the modulation is BFSK, the frequency spectrum shows two peaks at the two modulating frequencies, likewise; for 4-FSK the spectrum possesses four peaks, and generally for MFSK modulation, the spectrum has M peaks. Using this criterion, the author in [19] calculated the ratio between the second peak and the third peak. If the ratio is greater than one (or some specified threshold), then the third peak has very small a value compared to the second peak; in other words, there are only two peaks in the spectrum, and the modulation type is identified as BPSK. If the ratio is smaller than the specified threshold, the modulation type is definitely not BPSK. Another test is done between the fourth and the fifth peaks to recognize if the modulation type is 4-FSK or not, and so on. The ratio will always be unity for non-FSK modulations. This hierarchical approach fails if the signal contains any impulsive noise component other than AWGN [20].

3.3.2. Wavelet Transform.

The Fourier transform is the main analytical method used to study different signals in the frequency domain. However, the Fourier transform provides information about the spectrum of a signal assuming that the signal is stationary and its spectrum is time invariant. For non-stationary signals, the wavelet transform is introduced as a general solution to investigate the signal in both time and frequency domains. Moreover, the wavelet transform has a strong advantage of reducing the effect of noise on the transmitted signals.

In order to calculate the wavelet transform of a signal, the signal is processed by a group of high pass and low pass filters. The high pass filters allow high frequency components of the signal to pass (the detailed components of the signal). In contrast, the low pass filters suppress high frequency components of the signal and allow low frequencies to pass (an approximation of the signal) [21].

Figure 7 shows the wavelet decomposition method. It can be seen that the re-

ceived signal is first applied to a low pass filter (LPF) and high pass filter (HPF) resulting in an approximation and detailed coefficients of the signal, respectively. In the next level, the output of each filter is again applied to an HPF and an LPF with different cut-off frequencies from those used in the first stage. The result is another approximation and detailed coefficients of the signal. The final outputs of the 3-level wavelet transform are represented by the sequences D1, D2, D3, and A3 [22]. Digital modulated signals are band-limited signals that have different details at the same scale. Classification features can be extracted from the decomposed components and applied to a classifier to discriminate between the different modulation levels [23].

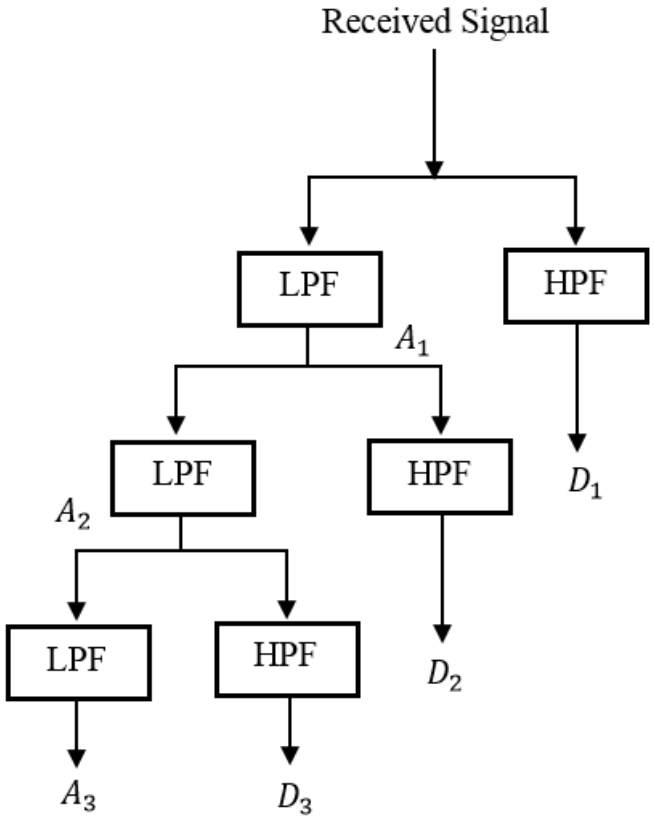


Figure 7: Wavelet decomposition.

3.3.3. Clustering Algorithms.

Clustering algorithms are used to group the received symbols of an unknown signal into clusters. The number of these clusters as well as the location of their cen-

troids can be used to identify the modulation type of the received signal. For example, a 4-QAM modulation has four different states. If the received signal is transmitted through a noisy channel, then the received symbols will reach the receiver scattered around their original positions. By using clustering algorithms, the scattered symbols are grouped based on their distribution resulting in four different clusters, which indicates that the modulation type is 4-QAM. Many clustering algorithms are suggested in the literature, such as k-means, fuzzy c-means, and subtractive clustering. For instance, subtractive clustering is used in [24] to classify different modulation types. Firstly, the likelihood of each received symbol to be a center of a cluster is calculated. This is usually done by measuring the Euclidean distances between each received symbol and all other received symbols in the constellation. The received symbol with the highest likelihood is selected as the center of the first cluster. Next, the likelihood of the remaining points to be a new center of a cluster is recalculated. This time, the likelihoods of the neighboring points to the first centroid are minimized. This is usually done to avoid selecting a very close point to the first centroid, since there is a high probability that the neighboring points belong to the same cluster. Again, the second centroid is selected. This process is continued until all the remaining points have no potential to be a centroid. The result is a final number of clusters with their centroids. Based on this information, the modulation type of the received signal is determined.

3.3.4. High Order Moments (HOMs) and High Order Cumulants (HOCs).

A signal's moments are the expected value of the signal raised to a power determined by the order of each moment. The mean of a signal is the first-order moment while the second-order moment usually indicates the power of the signal. For a complex-valued stationary random process, y_n, the p^{th} order moment is defined as:

$$M_{pq} = E\left[\mathbf{y}^{p-q}(\mathbf{y}^*)^q\right] \tag{21}$$

where $\mathbf{y}^*$ is the complex conjugate of $\mathbf{y}$, and q is the power of the conjugate signal $\mathbf{y}^*$. High-order cumulants are considered as another method to calculate a signal's statistics. In order to introduce signal cumulants, we first define the first characteristic function of any random variable x as [20]:

$$\Phi(s) = \int_{-\infty}^{\infty} f(x)e^{sx}dx \tag{22}$$

And its second characteristic function is defined by:

$$\varphi(s) = ln(\Phi(s)) \tag{23}$$

Signal cumulants are the derivatives of the second characteristic function around the origin:

$$\lambda_p = \frac{d^p \varphi(0)}{ds^p} \tag{24}$$

where p indicates the order of the cumulant.Generally, HOCs are expressed as functions of a signal's High Order Moments (HOMs) [20]. Typically, HOMs are calculated for zero mean signals. Thus, the mean is subtracted from the received data symbols as follows:

$$\tilde{y}_n = y_n - \frac{1}{N}\sum_{n=1}^{N} y_n \tag{25}$$

where N is the total number of symbols. Table 1 shows the relationship between some of the commonly used HOCs to the HOMs.

Table 1: High order cumulants

HOCs		HOMs Expression
Second-Order Cumulants	C_{20}	M_{20}
	C_{21}	M_{21}
Fourth-Order Cumulants	C_{40}	$M_{40} - 3M_{20}^2$
	C_{41}	$M_{40} - 3M_{20}M_{21}$
	C_{42}	$M_{42} - \lvert M_{20}\rvert^2 - 2M_{21}^2$
Sixth-Order Cumulants	C_{60}	$M_{60} - 15M_{20}M_{40} + 30M_{20}^3$
	C_{61}	$M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} + 30M_{20}^2M_{21}$
	C_{62}	$M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} - M_{22}M_{40} + 6M_{20}^2M_{22} + 24M_{21}^2M_{20}$
	C_{63}	$M_{63} - 9M_{21}M_{42} + 12M_{21}^3 - 3M_{20}M_{43} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22}$

It is noticed that the magnitude of the sixth and fourth-order cumulants is greater than the magnitude of second-order cumulants. This results in a different range of

values for the different HOC orders. Therefore, feature normalization is required in order for machine learning algorithms to work properly. For normalization, HOCs are rescaled as described in Table 2, with each cumulant raised to the power $\frac{2}{r}$, where r is the order of the cumulant [20].

Table 2: Normalized high order cumulants

Normalized HOCs	Normalization
$\hat{C}_{20}$	C_{20}
$\hat{C}_{21}$	C_{21}
$\hat{C}_{40}$	$C_{40}^{\frac{1}{2}}$
$\hat{C}_{41}$	$C_{41}^{\frac{1}{2}}$
$\hat{C}_{42}$	$C_{42}^{\frac{1}{2}}$
$\hat{C}_{60}$	$C_{60}^{\frac{1}{3}}$
$\hat{C}_{61}$	$C_{61}^{\frac{1}{3}}$
$\hat{C}_{62}$	$C_{62}^{\frac{1}{3}}$
$\hat{C}_{63}$	$C_{63}^{\frac{1}{3}}$

For more simplicity, the magnitude of the cumulants are used instead of their complex values. This step has a great advantage in reducing the processing time in the training stage because classifier's weights are real-valued in this case instead of being complex. Another major advantage is reduced vulnerability to shifts in the constellation, as the phase shift does not affect the magnitude of the cumulants though it may affect their imaginary part. Finally, some HOCs are more useful in separating a particular group of modulation types than others. Deciding which HOCs to use in each case is another challenge. The values of the normalized HOCs for the six modulation types BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM, at different SNR values, is presented in Appendix A.

The values of the fourth-order cumulants for Pulse Amplitude Modulation (PAM), Phase Shift Keying (PSK) and Quadrature Amplitude Modulation (QAM) under a noiseless environment are presented in [25]. It is shown that HOCs have different values for

different modulation schemes. For example, C_{20} is found to be zero for all types of QAM, PSK, and PAM, except 8-PAM. Based on this observation, C_{20} is calculated for every received signal and if C_{20} is not zero then the signal is classified as 8-PAM. On the other hand, if C_{20} is zero then it could be any of the other candidates. Finally, the modulation identity of the received signal is determined by comparing the values of the fourth-order cumulants to predefined thresholds [26]. The previous hierarchical approach is very common in classifying different modulated signals, since it is effective and has low computational complexity. Mixed-order moments are used in [27] to discriminate between different levels of QAM modulation. Two new parameters, V_{20} and V_{30}, are defined as:

$$V_{20} = \frac{M_{42}(s)}{{M_{21}}^2(s)} \tag{26}$$

$$V_{30} = \frac{M_{63}(s)}{{M_{21}}^3(s)} \tag{27}$$

These parameters are extracted from any received signal and projected using the Fisher criterion into one dimensional space. The modulation type is determined by comparing the projected value to some thresholds.

3.4. Optimization Techniques

In automatic modulation classification, optimization methods are used to find the best features from a group of features or sometimes to generate and test new features. Using optimization techniques has a great impact on increasing modulation classification accuracy. In each modulation classification problem, different classification features are calculated from the received signal and used to determine its modulation type. These features have different discriminant powers based on the modulation classification problem at hand. Optimization techniques are used to select the best features that impact the classification accuracy and remove the redundant features. The final output of the optimization step is a feature vector that has a fewer number of features compared to the original feature vector, as well as the same accuracy or sometimes better compared to the classification accuracy achieved using the original set of features. In general, heuristic optimization has several advantages such as:

- Insensitivity to the problem size.
- Providing a global solution.
- Providing a number of best solutions.

However, heuristic optimization techniques have disadvantages such as:

- Complexity.
- Long running time due to their iterative approach.

In the next section, two optimization techniques are introduced namely, genetic algorithms and particle swarm optimization.

3.4.1. Genetic Programming.

A genetic algorithm (GA) is a heuristic search algorithm invented by John Holland in 1970. GA applies the idea of natural selection and genetics on computer optimizations problems. GA is superior in problems where there is a large number of possible solutions and the search space has many hills and valleys. By conducting a parallel search and using the history of the previous solutions, GA is capable of finding the global optimum solution and oversees local optimum solutions that may not be easily avoided by most other optimization techniques. GA is widely used on different optimization problems when the objective is to find the best candidates from a certain population that lead to the optimal solution. The selection criteria are employed using operators such as mutations and recombinations.

In automatic modulation classification, several features are proposed to distinguish between different modulation types. These features differ in their classification power. GA is used to select the best features and discard the redundant ones to avoid using a large number of redundant features that might result in reducing the classification accuracy (this phenomenon is referred to as the curse of dimensionality). The implementation of GA requires:

- Fitness function: the objective function that should be maximized or minimized by the genetic programming algorithm. For automatic modulation classification, it can be chosen as the classification accuracy.
- Initial population: genetic programming needs initial values that represents possible solutions to the problem (classification features for the case of automatic mod-

ulation classification). The initial population is selected randomly from the group of possible solutions in order to cover all the possible solutions. These solutions are evaluated by the fitness function to determine the best solutions among the initial population. The best solutions are then used in the next iteration whereas the unselected solutions are discarded.

- Crossover and mutations: the surviving population, usually referred to as parents, are used to generate a new group of possible solutions known as children. Children are generated from their original parents by a process called crossover. In the crossover process, different functions such as plus, minus, times, reciprocal, negator, abs, sqrt, sin, cos, tan, asin, acos, and tanh [28] are used to combine the different parents and produce new children with mutual features from their original parents. Moreover, another process called mutation is used to slightly change the produced child in order to keep the randomness behavior in the generation of the new population.
- Stopping criteria: after generating the new population, children solutions are evaluated using the predefined fitness function, and the best children among the new solutions are kept for the next iteration to represent the new parents. Usually the fitness of solutions in the second iteration is higher compared to the fitness of solutions in the first iteration. The process of combing the features and evaluating the solutions continues for several iterations until a certain condition is satisfied; either the number of iterations reaches its maximum, or the fitness of the new generations has no significant addition compared to the previous solutions.

3.4.2. Particle Swarm Optimization.

The Particle Swarm Optimization (PSO) algorithm is a robust optimization technique developed in 1995 [29]. PSO imitates the behavior of flocks of birds, swarms of insects, or schools of fish when traveling together as a group to reach their final destination. In PSO, particles are flying through a multidimensional search space where each particle adjusts its position according to what it experiences along with feedback from all the neighboring particles. In PSO, all particle positions are initiated randomly and evaluated to compute a predefined fitness function.

For automatic modulation classification, classification features represent the particles in the swarm, and the fitness function is defined as the achieved classification accuracy when using these features. In the first iteration, a group of solutions (particles) is initiated randomly, where each particle is a vector of features. The particle with the highest classification accuracy is referred to as the local best, whereas the particle that leads to the best classification accuracy in all the iterations is referred to as the global best [30]. Features in each particle are changed based on its previous experience and the experience of the entire swarm in order to increase its fitness function (classification accuracy) in the next iteration. By the end of this optimization technique, the output is a group of particles (vectors of features) that lead to the best classification accuracy. These features are then used in the automatic modulation problem instead of the original set of features.

Chapter 4: Pattern Recognition Techniques

"Pattern recognition is the scientific discipline whose goal is the classification of objects into a number of categories or classes " [31]. Pattern recognition has many applications in different fields. For instance, *Machine Vision* is used to analyze captured images and determine their type, for example in the manufacturing industry to determine whether the product is defective or not. *Biomedical Applications* help physicians in diagnosing the different types of diseases, and *Character recognition* is used to identify written texts like letters and numbers that can be used to identify signatures in banks, or in supermarkets price checkers.

In pattern recognition, the identity of different patterns is determined from their distinguishing features. The accuracy of this approach depends on the discriminating power of the selected features (classification features) and the capability of the applied classifier. Pattern recognition can be divided into three main categories: *Supervised Learning Algorithms* where labeled data from known classes are used to train the machine learning classifier; *Unsupervised Learning Algorithms* in which the class identity of the training data is not available at the classifier, which is usually harder to solve; and finally *Semi-Supervised Learning Algorithms*, which is a hybrid classification type when part of the training data is labeled and the other part is not.

Any pattern recognition system typically consists of a number of stages as shown in Figure 8. Input data is usually raw data like images (in image processing problems), human voice (in speech recognition problems) or modulated received signals (in automatic modulation classification problems). The preprocessing stage is used to clean the patterns of interest, and usually noise filtering, data normalization, and data smoothing are done in this phase [32]. After that, the selected discriminant features are extracted from the processed input data in order to form the actual input to the machine learning classifier. An example of feature extraction is calculating the signal amplitude and phase (in automatic modulation classification problems). The last stage in any pattern recognition system is choosing the classifier model by which the class identity is determined. Different types of classifiers are examined in this thesis such as the K-Nearest Neighbor Classifier, Naïve Bayes Classifier, Artificial Neural Network

Classifier, Support Vector machine Classifier, and Polynomial Classifier.

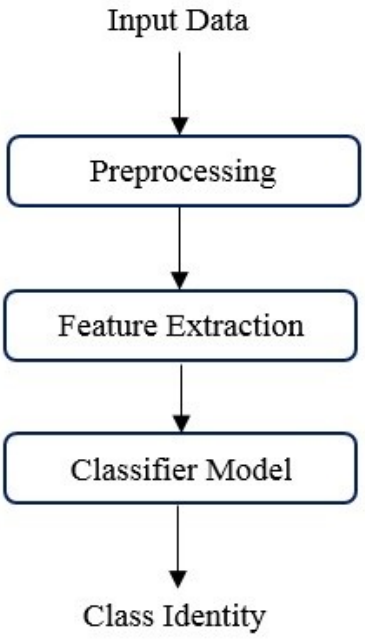


Figure 8: Pattern recognition system.

4.1. K-Nearest Neighbor Classifier

The K-Nearest Neighbor Classifier (KNN) is one of the oldest and simplest pattern recognition classifiers. It is a non-parametric technique that requires little or no information about the data distribution [33]. In a KNN classifier, labeled training data is used as a reference and based on their distribution the identity of an unknown tested sample is determined. A KNN classifier calculates the distances between the tested sample and all the labeled training samples; then the identity of the tested sample is determined by the class of its majority neighbors. Different parameters may impact the performance of a KNN classifier such as:

- Number of neighbors (K): the number of neighbors used by a KNN classifier to determine the class of an unlabeled sample. The selection of an appropriate value of K is an important parameter to obtain high classification accuracy and maintain system simplicity. However, the best value of K differs from one application to another, and its selection depends on the problem at hand.

- Distance measures: different types of distances can be used as a measure when using a KNN classifier. For a classification problem with *n-dimensional* feature vectors, the distance between a test sample $\mathbf{x}$ and a training sample $\mathbf{y}$ is defined as follows:

    1. The Euclidean distance: the most commonly used measure (the default in many programming languages), described as:

$$D(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{28}$$

    2. City block distance: Also known as Manhattan distance, city block distance is very similar to the Euclidian distance, but has less sensitivity to large differences in one dimension (each dimension represent a feature), since the square value of the distances is not calculated in this case:

$$D(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} |x_i - y_i| \tag{29}$$

    3. Cosine distance: Is a measure of similarity between two points calculated by:

$$D(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} \tag{30}$$

    where $(.)^T$ indicates the transpose operation.

    4. Correlation distance: Is another similarity measure that takes values between zero to one. A correlation distance of one indicates that the samples belong to the same class:

$$D(\mathbf{x},\mathbf{y}) = \frac{cov(\mathbf{x},\mathbf{y})}{\sqrt{Var(\mathbf{x})Var(\mathbf{y})}} \tag{31}$$

where $cov(\mathbf{x},\mathbf{y})$ is the covariance matrix between vector $\mathbf{x}$ and vector $\mathbf{y}$, and $Var(\mathbf{x})$ and $Var(\mathbf{y})$ are the variance in vector $\mathbf{x}$ and vector $\mathbf{y}$, respectively.

An example of a KNN algorithm using Euclidean distance is showed in Figure 9. Two features are extracted from each sample and are mapped into two dimensional space. For a tested sample represented by the cross, its distance from all the training samples is calculated, and for a KNN classifier with $K = 5$. Three of the closest samples are from class B while only two samples are from class A. Accordingly, the class of the test sample is assigned as B.

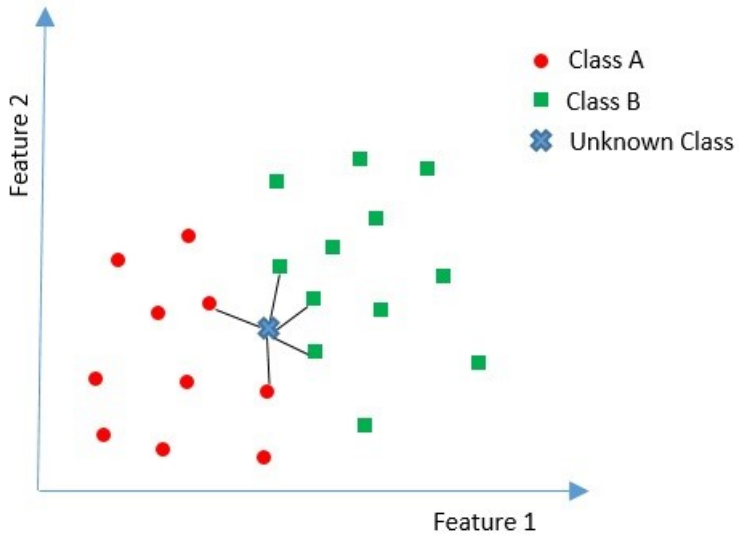


Figure 9: KNN classifier.

4.2. Naïve Bayes Classifier

This is one of the simplest classifiers that is originally based on the Bayesian theory of probability. It makes the naïve assumption about the different features in the classification problem that they are totally independent. A naïve Bayes classifier requires a small set of training data to estimate the data distribution (mean and variance). Since no dependency between the features is assumed, only the variance is calculated (not the covariance matrix) as in the case of dependent features. In spite of its simplicity, the naïve Bayes classifier was tested for a variety of real problems and showed a very good performance even in some cases when the different attributes are not independent. It also has an advantage when the sample size is small [34]. The conditional probabil-

ity of a tested sample to be from class C, given the set of features $[x_1, x_2, x_3, ...x_n]$, is calculated by the naïve Bayes classifier as:

$$P(C/x_1, ..., x_n) = \frac{P(C)P(x_1, ..., x_n/C)}{P(x_1, ..., x_n)} \tag{32}$$

where $P(C)$ is the prior probability of occurrence of class C, $P(x_1, ..., x_n)$ is the probability of having the feature vector $[x_1, x_2, x_3, ...x_n]$, and $P(x_1, ..., x_n/C)$ is the probability of having the feature vector $[x_1, x_2, x_3, ...x_n]$ given that the class of the signal is C.

After we apply the naïve assumption that all the features are independent:

$$P(C/x_1, ..., x_n) = \frac{\prod_{i=1}^{n} P(C)P(x_i/C)}{P(x_1, ..., x_n)} \tag{33}$$

Since $P(x_1, ..., x_n)$ is equal for all the classes:

$$P(C/x_1, ..., x_n) \; \alpha \; P(C) \prod_{i=1}^{n} P(x_i/C) \tag{34}$$

Thus, the selected class is the one that has the largest value:

$$\hat{C} = \arg\max_{C} \left\{ P(C) \prod_{i=1}^{n} P(x_i/C) \right\} \tag{35}$$

The naïve Bayes classifier is very simple and fast in delivering the final decision. Therefore, it is suitable for real-time applications.

4.3. Artificial Neural Networks

Artificial Neural Networks (ANNs) or Neural Networks (NNs) are mathematical models inspired by biological nervous systems such as the human brain. They were first introduced in 1943 as a computational model called the threshold model [35]. Researchers continued the work on artificial intelligence and created different models of neural networks that exist today. Neural networks have the ability to estimate the non-linear relationship between a system's inputs and outputs. Moreover, it can also find the patterns and trends in complicated data that might be difficult to extract using other

techniques. NNs require a training stage, wherein the system is provided with different patterns from different data sets in order to estimate a mathematical model for the problem. It is followed by a testing stage in which the performance of the classifier is evaluated and tested.

NNs can be categorized into supervised and unsupervised based on their learning method. They can also be classified as a single layer or multilayer based on their connection structure. In single layer construction, inputs (which represent the input features and a bias in our automatic modulation classification problem), are connected to each other and multiplied by weights to give the final output. The multilayer structure is the most commonly used configuration. It has three main layers an input layer (that accepts features and a bias value as inputs) and activates the significant inputs, a hidden layer (a combination of the activated input features multiplied by their corresponding weights represents the input to this layer), again another activation functions are used in this layer to keep the significant combinations. Finally the output layer, which is the final output from the neural network, is used to determine the class of the unknown modulation type. Figure 10 shows an example of 3 layered NNs with M different classes.

The iutput to each neuron in the hidden layer is:

$$g_j = \sum_{i=1}^{n} w_{ji} * x_i + w_{j0} \tag{36}$$

where n is number of the activated neurons in the input layer, w_{ji} is the weight of the link joining neuron i in the input layer to neuron j in the hidden layer, and w_{j0} is the weight of the link connecting the bias input to neuron j. Figure 11 shows neuron j in the hidden layer, where an activation function is applied on the input g_j to define the output z_j. Different activation functions are used in NNs, however, one of the simplest functions is defined as:

$$z_j = f\left(g_j\right) = \begin{cases} 1 & if \quad g_j \geq 0 \\ -1 & if \quad g_j < 0 \end{cases} \tag{37}$$

Finally, in the output layer, C outputs are calculated, where each output repre-

sents a different class. Figure 12 shows neuron k in the output layer, where its input is defined as:

$$m_k = \sum_{j=1}^{b} w_{kj} * z_j + w_{k0} \tag{38}$$

where b is number of the activated neurons in the hidden layer. Accordingly, the final output of an NN classifier is expressed as:

$$y_k = f(m_k) = \begin{cases} 1 & if \quad m_k \geq 0 \\ -1 & if \quad m_k < 0 \end{cases} \tag{39}$$

where y_k is a nonlinear function of the input features x_i. Moreover, the DC input bias allows the final classifier model to be a non-homogeneous function of the original input features (most of the practical discriminant functions have a DC constant term).

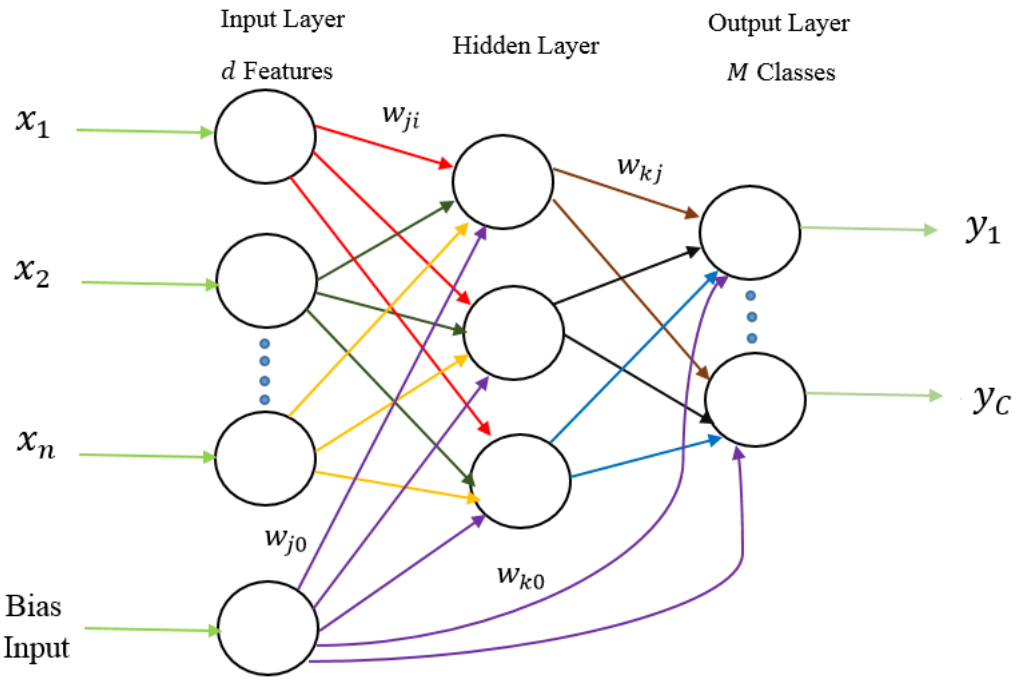


Figure 10: Three layer neural networks.

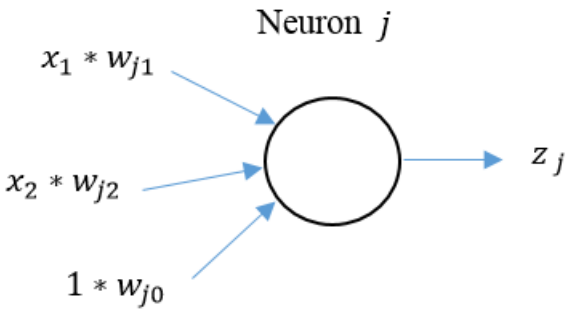


Figure 11: Neuron j in the hidden layer.

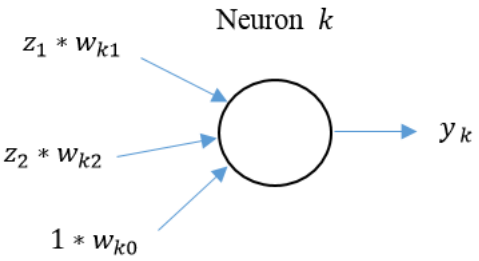


Figure 12: Neuron k in the output layer.

4.4. Support Vector Machines

Support Vector Machines (SVMs) are a powerful supervised machine learning algorithm introduced in 1995 [36]. An SVM is essentially designed for binary classification problems where the decision is either one class or an other. SVM can easily solve classification problems when the two classes are non-linearly separable by simply mapping the input set of features into a higher dimensional space in which the given classes are linearly separated.

In the training phase, the classifier estimates a separating hyper-plane between the two classes in the new dimensional space. Usually there is more than one hyper-plane that satisfies the criterion. However, the SVM classifier chooses the best one that yields to the maximum distance between the support vectors (support vectors are the closest samples from the two classes to the separating hyper-plane). In the testing phase, an unlabeled testing sample is mapped to the same high dimensional space used in calculating the model. And its identity is determined based on its position from the boundary. Figure 13 shows a binary classification problem in a two-dimensional space; the separating boundary is perpendicular to the line that maximizes the distance

between the closest support vectors from different classes.

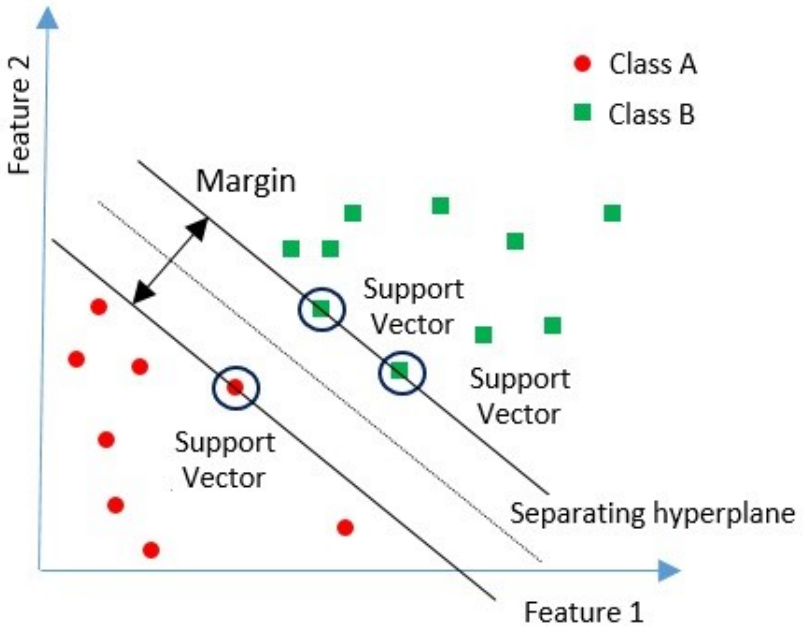


Figure 13: SVM classifier.

Chapter 5: Proposed Hierarchical Automatic Modulation Classification Scheme

The objective of this work is to develop an AMC scheme that would allow the receiver to decide which of the modulation types is used by the transmitter to send a given block of data. AMC is treated as a pattern recognition problem, where features are extracted from the received signals and used in a classifier to decide upon the modulation level. The features used in this work are the HOCs of the received signal and the classifier used is a *Hierarchical Polynomial* (HP) classifier. It is noted that the performance of any pattern recognition scheme depends on the selected features applied to the classifier and the capability of the classifier itself. In order to improve the classification accuracy of the HP classifier, an SNR estimation system is proposed, where the estimated SNR is fed to the classifier and used to optimize the classifier's weights in order to improve the overall classification accuracy. Moreover, feature selection system based on stepwise regression is proposed, where the important features for the modulation classification are identified and the redundant features are discarded. The result is a simplified modulation classifier with optimized performance.

5.1. Signal Model

In this work, the signals are assumed to be transmitted over a flat fading channel, which is a valid and rational assumption for narrow band systems. For example, an OFDM system modulates a group of subcarriers, with each subcarrier carries data with a low symbol rate. Thus, the reflected signals from the different paths arrive within a short time period smaller than the symbol duration. Therefore, there is no inter-symbol interference (ISI) between the different consecutive transmitted symbols. The baseband discrete-time received signal contaminated by Additive White Gaussian Noise (AWGN) in a flat fading environment can be expressed as:

$$y_n = h_n x_n + w_n \quad n = 1, ..., N \tag{40}$$

where x_n is the discrete time transmitted signal, w_n is the AWGN process with zero mean and two-sided power spectral density $\frac{N_0}{2}$, h_n is the complex valued channel gain assumed to follow a Gaussian distribution, and N is the number of transmitted symbols. The transmitted signal x_n is selected to be one of L possible modulation types. In this work we consider BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM or 256-QAM modulation types ($L = 6$). Furthermore, the transmitted power is normalized to unity for all the modulation types.

5.2. Hierarchical Polynomial Classifier

The Weierstrass approximation theorem states that "every real-valued continuous function on a finite closed interval [a, b] can be uniformly approximated by polynomials with real coefficients" and "every complex-valued continuous function on a finite closed interval [a, b] can be uniformly approximated by polynomials with complex coefficients" [37]. According to this theorem, a polynomial classifier can be used to approximate the nonlinear boundaries between different classes. A polynomial classifier is a machine learning algorithm that expands the original set of features in a given space to a higher dimensional space in which the different classes are linearly separated. Due to its simplicity and high accuracy, it has applications in a variety of fields like speech recognition [38], cognitive radio systems [39], and biomedicine [40].

Figure 14 shows a binary classification example in the original feature's space d_1 and d_2. It can be noticed that the two classes are nonlinearly separable and can only be separated by a quadratic function. However, the new set of features, x_1, x_2, and x_3, derived from an expansion of d_1 and d_2, can linearly separate the two classes in a higher dimensional space (3 dimensions in this case), where $x_1 = d_1^2$, $x_2 = d_2^2$, and $x_3 = \sqrt{2}d_1d_2$.

Similar to any supervised learning algorithm, a polynomial classifier has two main stages, as shown in Figure 15. First, there is a *training* stage in which features from labeled training signals are used in calculating the classifier weights, and then there is a *testing* stage where unlabeled signals are applied to the classifier to identify the classes of the signal.

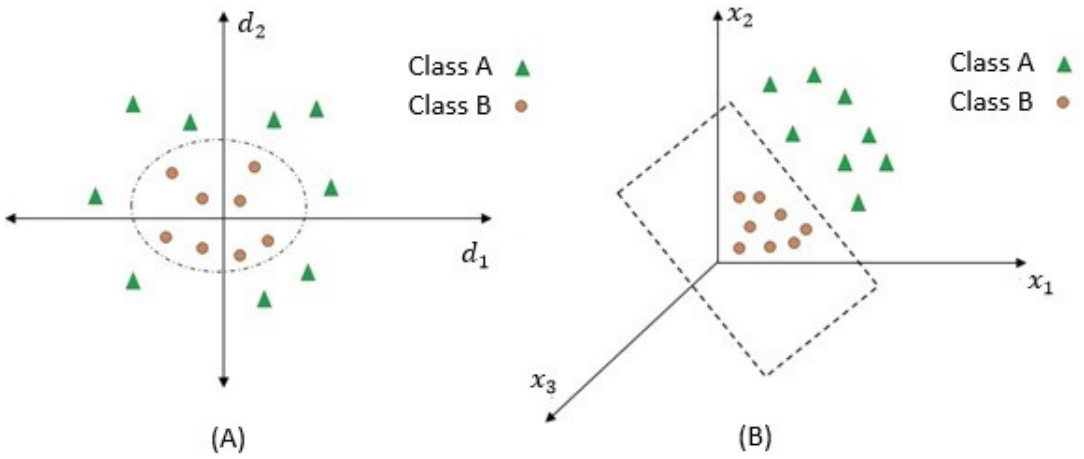


Figure 14: Classification features in the original space and high dimensional space.

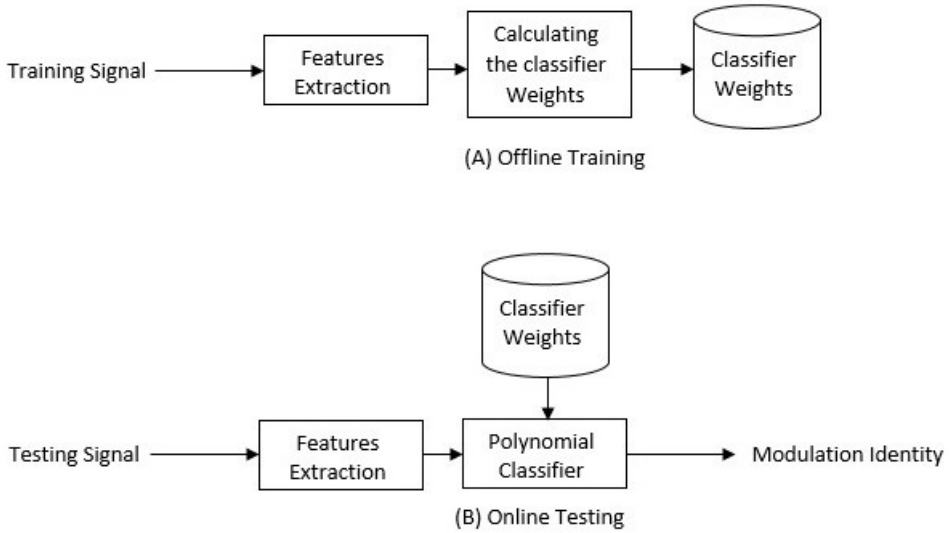


Figure 15: Training and testing stages.

The proposed HP classifier is shown in Figure 16, where the classification of the six ($L = 6$) modulation types (BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM) is done in hierarchical binary-classification stages. In each stage, modulations having features of similar values are clustered in one class and the rest are placed in a second class. For example, the received signal is firstly classified to be either PSK or QAM. In the next stage, if the signal is PSK, two new classes are introduced, BPSK as a class and QPSK and 8-PSK as another class. If the signal is classified as BPSK, the classification procedure is completed; otherwise, another *binary* classification between QPSK and 8-PSK is introduced. The same principle applies for the QAM types as shown in Figure 16. This process helps in optimizing the weights in each binary

classification stage leading to an overall improvement in the classification accuracy. In contrast, a conventional polynomial classifier used in [9] has less classification accuracy, especially for high values of L (more considered modulation types). For example, it is much easier to identify 64-QAM when the modulation is either 64-QAM or 256-QAM, compared to a case when the modulation is either of BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, or 256-QAM.

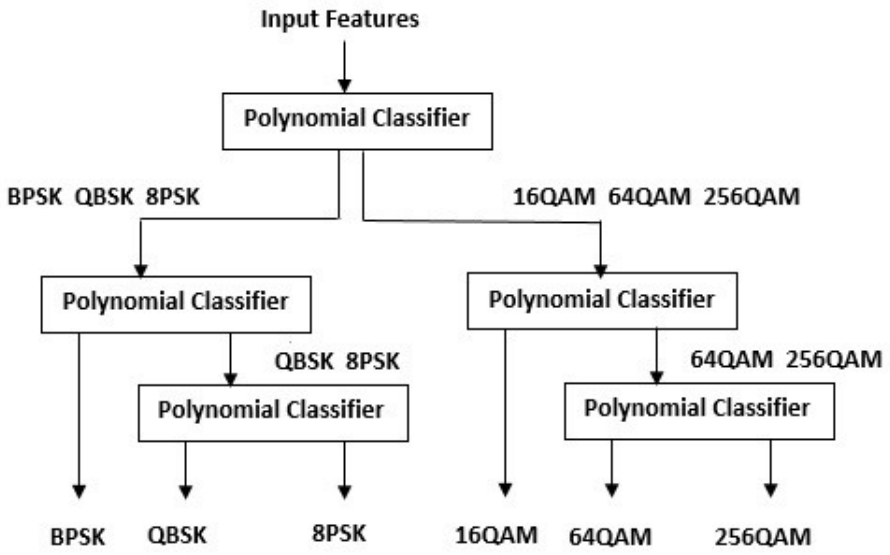


Figure 16: Hierarchical polynomial classifier.

5.2.1. Training Stage.

HOC features are calculated from each received signal with known classes. The obtained features are expanded into a higher dimensional space using a polynomial classifier in order to produce more features and allow for easier separation of the classes. The order of the polynomial classifier determines the dimensionality of the space. Although higher order polynomial classifiers could be used, for implementation simplicity, a second-order polynomial classifier is normally implemented, where the new set of features is the original features plus their products and squared values. For instance, for an input feature vector $\mathbf{d}$ composed of normalized HOCs defined as $\mathbf{d} = [d_1, d_2, \ldots, d_M]$, the expanded feature vector $\mathbf{p}$ is expressed as:

$$\mathbf{p} = \begin{bmatrix} 1, & d_1, & d_2, & \dots, & d_M, \\ d_1 \times d_2, & d_1 \times d_3, & \dots, & d_1 \times d_M, \\ d_2 \times d_3, & d_2 \times d_4, & \dots, & d_2 \times d_M, \\ \vdots & & & \vdots \\ d_{M-1} \times d_M, & d_1^2, & d_2^2, & \dots, & d_M^2 \end{bmatrix} \tag{41}$$

where M is the number of HOCs used as input features. For simplicity, vector $\mathbf{p}$ can be written as:

$$\mathbf{p} = [p_1, \ p_2, \ \cdots \ , p_n] \tag{42}$$

where n is the length of the higher dimensional feature vector.

To explain how the polynomial classifier works, let us consider a case of two modulation types, for example BPSK and QPSK (i.e. $L = 2$). In order to train the polynomial classifier, K training signals are transmitted from each of the BPSK and QPSK modulations. HOCs are extracted from each signal and the corresponding $\mathbf{p}$ vector is calculated using (41). After that, a new matrix $\mathbf{v}_l$ that has all the expanded feature vectors for modulation type l is defined as:

$$\mathbf{v}_l = \begin{bmatrix} \mathbf{p}_1^{(l)} \\ \mathbf{p}_2^{(l)} \\ \vdots \\ \mathbf{p}_K^{(l)} \end{bmatrix} \qquad l = 1, 2 \tag{43}$$

By rearranging both $\mathbf{v}_1$ for the BPSK modulation and $\mathbf{v}_2$ for the QPSK modulation in a new matrix, $\mathbf{V}$, we get:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{44}$$

The next step is to calculate the optimum classifier weights, $\mathbf{w}_1^{opt}$ for BPSK and $\mathbf{w}_2^{opt}$ for QPSK. In order to do so, the square of the second norm is minimized as:

$$\mathbf{w}_l^{opt} = \arg\min_{\mathbf{w}_l} \|\mathbf{V}\mathbf{w}_l - \mathbf{t}_l\|^2 \tag{45}$$

where $\mathbf{t}_l$ is the target matrix defined as:

$$\mathbf{t}_l = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \end{bmatrix}, when\ calculating\ \mathbf{w}_1^{opt} \tag{46}$$

and

$$\mathbf{t}_l = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, when\ calculating\ \mathbf{w}_2^{opt} \tag{47}$$

with

$$\mathbf{0} = \lfloor 0,0,...,0 \rfloor_{1\times K}^{\mathbf{T}} \tag{48}$$

$$\mathbf{1} = \lfloor 1,1,...,1 \rfloor_{1\times K}^{\mathbf{T}} \tag{49}$$

Equation (45) can be simply written as:

$$\mathbf{V}^{\mathbf{T}}\mathbf{V}\mathbf{w}_l = \mathbf{V}^{\mathbf{T}}\mathbf{t}_l \tag{50}$$

$$\sum_{m=1}^{2} \mathbf{V}_m^{\mathbf{T}}\mathbf{V}_m\mathbf{w}_l = \mathbf{V}^{\mathbf{T}}\mathbf{t}_l \tag{51}$$

However, by defining:

$$\mathbf{Y} = \sum_{m=1}^{2} \mathbf{V}_m^{\mathbf{T}}\mathbf{V}_m \tag{52}$$

Then:

$$\mathbf{w}_l^{opt} = \mathbf{Y}^{-1}\mathbf{v}_l \tag{53}$$

where $\mathbf{Y}^{-1}$ is the inverse of matrix $\mathbf{Y}$ [41]. Finally, after finding $\mathbf{w}_1^{opt}$ and $\mathbf{w}_2^{opt}$, the classifier is ready for the testing stage.

5.2.2. Testing Stage.

In the testing stage, the objective is to find the identity of an unlabeled modulated signal, either BPSK or QPSK. To classify the modulation type, the HOC feature vector

d is first extracted from the received signal and the expanded vector **p** is calculated using the second-order expansion. Then, vector **p** is multiplied by the classifier weights $\mathbf{w}_1^{opt}$ and $\mathbf{w}_2^{opt}$ obtained during the training stage to give two scores s_1 and s_2, respectively. These scores represent the new super features for the polynomial classifier, and based on their values the modulated signal identity is determined. Ideally, the weights are optimized during the training stage to give $s_1 = 1$ and $s_2 = 0$ if the modulation is type one (BPSK), and to give $s_1 = 0$ and $s_2 = 1$ if the modulation is type two (QPSK). Since the received symbols are noisy, the decision is made based on the maximum values of s_1 and s_2, meaning that if $s_1 > s_2$ then the modulation is class one, and if $s_2 > s_1$ the modulation is class two. Hence,

$$Class\ \ identity\ \ l = \arg\max_{l} \{s_l\} \tag{54}$$

The previous example explained the modulation classification process for only two modulation types, BPSK and QPSK. In this work, we consider a more general classification problem among BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM. Using the same concept introduced before, we divide the modulation classification process into binary sub-classification stages. In the first stage, PSK signals are treated as one class and QAM signals as another class. In other words, we define a problem of two modulation groups, the PSK modulation group that has BPSK, QPSK and 8-PSK signals, and the QAM modulation group that has 16-QAM, 64-QAM and 256-QAM. K training signals form BPSK, QPSK and 8-PSK are used to calculated matrix $\mathbf{v}_1$, and K training signals form 16-QAM, 64-QAM and 256-QAM are used to calculated matrix $\mathbf{v}_2$. Following the same procedure, $\mathbf{w}_1^{opt}$ and $\mathbf{w}_2^{opt}$ are calculated in the training stage. In the testing stage, if $s_1 > s_2$, the received signal is PSK and if $s_2 > s_1$, the received signal is QAM. If the signal is decided to be PSK, another binary modulation classification problem is considered. This time, another polynomial classifier is used, where matrix $\mathbf{v}_1$ has the expanded feature vector of the BPSK modulation, and matrix $\mathbf{v}_2$ has the expanded feature vector of both QPSK and 8-PSK. Accordingly, in the testing stage, if $s_1 > s_2$, the received signal is BPSK, and if $s_2 > s_1$ then the received signal is either QPSK or 8-PSK. Finally, if $s_2 > s_1$ then another polynomial classifier is used to determine the modulation type identity as either QPSK or 8-PSK using the same illustrated

concept. The same idea is used if the signal is decided to be QAM, as illustrated in Figure 16.

5.3. Stepwise Regression

Solving the classification problem in higher dimensional space proved to be advantageous; however, the dimensionality of the final feature vector grows dramatically with the increase in the number of input features. For example, for two input features, the corresponding expanded vector will have five features, whereas for our scenario with nine input features, the final expanded feature vector has a total of fifty-four features. Moreover, some of the features in the new dimensional space are redundant; therefore, calculating them adds to the complexity of the problem. In fact, it may result in decreasing the classification accuracy due to the possibility of having an over-fitted model. However, in this section we propose a combined model based on polynomial classifiers and stepwise regression for feature selection.

A stepwise regression model is used to select the significant features from the expanded feature vector **p**. In order to do so, the algorithm starts by constructing various regression models, each one based on one feature from the expanded feature vector. For example, for a model based on the first feature p_1, the regression model is defined as:

$$\hat{Y} = B_0 + B_1 p_1 \tag{55}$$

where $\hat{Y}$ is the predicted output based on the fitted regression model. The regression coefficients $B_0, and B_1$ are calculated by assuming the input to the model to be p_1, and the output of the model is Y, where Y represents the modulation label. The p-value of each feature is then calculated and the lowest p-value indicates the most significant feature. The p-value is defined as the null-hypothesis test probability (the null-hypothesis in this case is the event that the feature has no statistical significance).

Typically, the p-value is calculated with a tolerance of α for adding the feature to the model, and a tolerance of β for removing the feature from the model. In the next iteration, the p-value is calculated for all the remaining features given that the best feature has already been selected from the first iteration. The output of this iteration

is the best two-feature model. In the third stage, another feature is added to the model and tested; moreover, the model is examined when removing the first selected feature and keeping the other two features. If the p-value of the first feature is less than β, the first feature is removed from the model. Finally by the end of all iterations, the output of the regression model is the best features for the classification problem. Algorthim 1 shows the stepwise regression process explaining both of the adding and the removing processes [42].

The significance of each regression model is tested using the partial F-statistic parameters [1] defined as:

$$F_j = \frac{SSR\left(B_j/B_0, B_1, ..., B_{j-1}, B_{j+1}, ..., B_k\right)}{MSE} \tag{56}$$

where the F-statistic parameters are calculated when adding the term p_j given that the coefficients $B_0, B_1, ..., B_{j-1}, B_{j+1}, ..., B_k$ are already in the model. MSE is the mean square error of the model and SSR is the sum of squares defined as:

$$SSR = \sum_{n=1}^{N} \left(\hat{Y}_t - \bar{Y}\right)^2 \tag{57}$$

where $\bar{Y}$ is mean of the outputs defined as:

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{58}$$

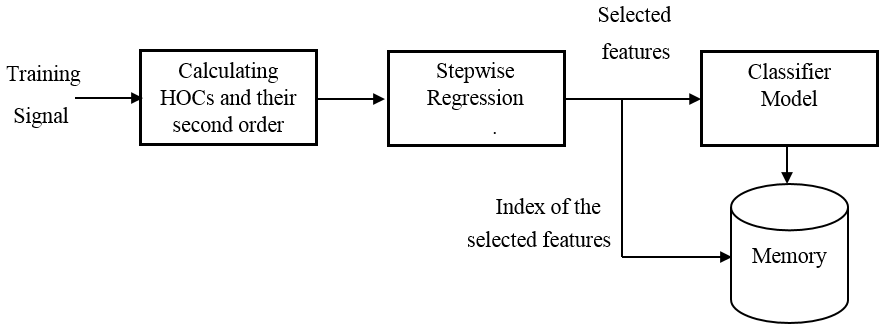
and n is the number of observations used in calculating the regression model.

Finally, the stepwise regression model is integrated with the polynomial classifiers. In the training stage, the second-order expansion is calculated for each feature vector and examined by the stepwise regression model. Therefore by the end of the regression process, only the significant features are used to train the classifier; thus the final classifier model is a function of the best calculated features. Using this approach, the stepwise regression process is required in the training stage only, and no heavy calculations are conducted in the testing stage. In fact, only the important features are extracted from the received signal in the testing stage, which reduces the complexity of the system in the real life scenario. Figure 17 shows the integrated model with both the

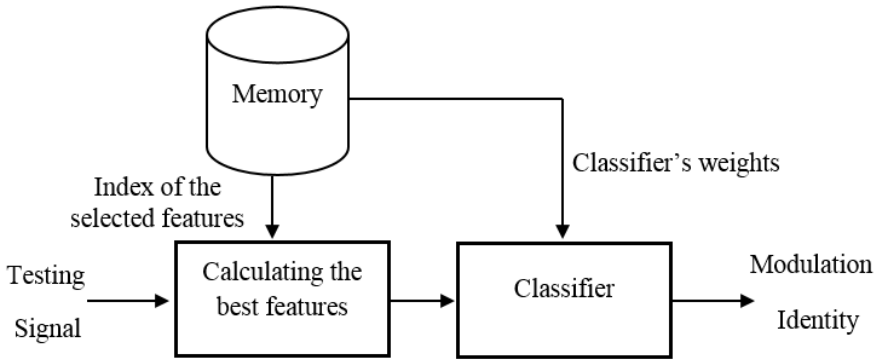
stepwise regression system and the polynomial classifier.

Algorithm 1 Stepwise regression

1: **procedure** START THE REGRESSION
2: *Calculate the P-Value*
3: **if** *Terms have* $P-Value < \alpha$ **then**
4: Add the term with the minimum P-Value
5: **else if** *Terms have* $P-Value > \beta$ **then**
6: Remove the term with the largest P-Value



(a) Training Stage



(b) Testing Stage

Figure 17: The proposed classifier using stepwise regression.

5.4. Signal-to-Noise Ratio Estimation

In this chapter, a new Signal-to-Noise Ratio (SNR) estimation method based on C_{20} and the naïve Bayes classifier is proposed. The SNR estimation system is then

integrated with the proposed automatic modulation classifier in order to improve the overall classification accuracy.

In general, received signals at different SNR values have different patterns, since the effect of noise in each case is different. Accordingly, some HOCs express some variations in their values at different SNRs. These variations lead to two different classifier training scenarios. In the first scenario, training signals at different SNR values are used to calculate average classifier weights, where the classifier is trained for different HOC patterns. As a result, the final classifier model can accurately identify signals at an unknown SNR. In contrast, the second training scenario intends to train the classifier for a specific pattern of HOCs, in other words at a specific SNR value; this leads to the best classification accuracy if the tested signals are at the same SNR value. However, the second scenario has a severe drop in classification accuracy if the received signal is at a different SNR value. Hence, for the second scenario to work, an accurate SNR estimation system is required.

In this section, we propose an SNR estimation algorithm that can be integrated with our modulation classifier. The power of the received signal (C_{20}) is used as an identification feature to estimate the SNR of the signal. The power of the received signal equals the power of the transmitted signal plus the power of the noise. Without loss of generality, we assume that the transmitted power is normalized to unity. Hence, C_{20} can accurately indicate the power of noise in the received signal, in other words its SNR value.

This approach is able to identify the SNR of the signal when there is a significant noise power added to the signal, (i.e. at low SNR values), but fails at high SNRs where the value of C_{20} is always close to unity. For that reason, we limit our estimator to SNR values in the range of 0 dB to 10 dB. We define 12 different classes where class one to eleven represent the SNR value 0 dB,1 dB,2 dB, ..., 10 dB, respectively, and the twelfth class represents any SNR value greater than 10 dB. A naïve Bayes classifier is used to decide on the SNR of the channel.

It is worth mentioning that the actual SNR of the signal is a continuous value and not discrete as we defined in our estimator. However, the estimated SNR is only used to select the corresponding classifier's weights. Hence, an approximated value is suf-

ficient to achieve a significant improvement in the modulation classification accuracy. Figure 18 shows the testing stage for the integrated SNR estimator and the proposed polynomial classifier system.

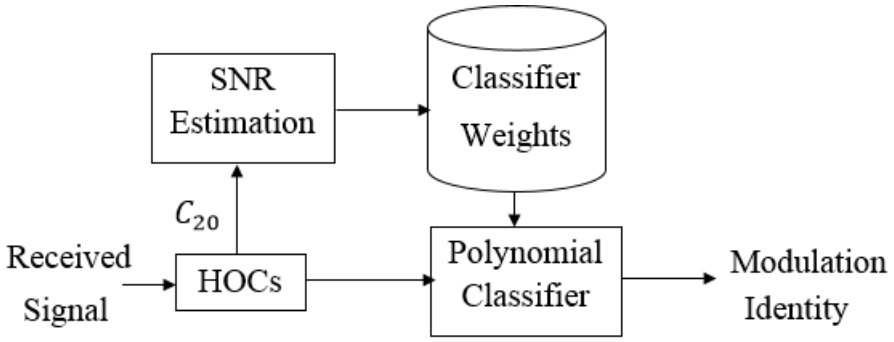


Figure 18: Modulation classification with SNR estimation (testing stage).

Chapter 6: Simulation Results

In this section, the performance of the proposed Hierarchical Polynomial (HP) classifier is examined under different channel conditions and is compared to other methods from the literature. First, the advantage of the proposed classifier over the traditional threshold-based method is investigated. A simulation involving the generation of 1000 different realizations of each of 16-QAM and 64-QAM signals is conducted at a 20 dB signal-to-noise (SNR), where each signal represents a block of $N = 2000$ symbols. The conventional method for classifying 16-QAM and 64-QAM signals uses the value of the fourth-order cumulant C_{42}, or the value of the sixth-order cumulant C_{63} as defined in [43]. Using a threshold to decide on the modulation type is not the optimal solution. Figure 19 shows that there are many signal misclassification errors when the classification is done based on a threshold value of C_{42} only or C_{63} only. On the other hand, it is clear from Figure 19 that the two modulation types can be easily classified using a line which is a function of C_{42}, C_{63} and a constant value. Figure 20 shows the two output scores when using the proposed polynomial classifier with the same input features C_{42} and C_{63}. In this scenario, when 16-QAM is transmitted (the first 1000 signals) then the first score s_1 is always greater than the second score s_2. The opposite happens when the transmitted signal is 64-QAM (signals numbered 1001 to 2000) where the second score s_2 is always greater than the first score s_1. Using the values of s_1 and s_2 results in a classification rate close to 100%.

6.1. Classification in AWGN Channels

All simulations in the next section are conducted for $N = 10000$ symbols. In Figure 21, a classification problem among BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, and 256-QAM is considered under AWGN channel conditions using 2000 realization for each modulation type. First, we have compared the performance of the proposed HP classifier with the *conventional polynomial* (CP) classifier, i.e. nonhierarchical, under the assumption that no SNR information is available at the receiver side. During the training stage, signals with different known SNR values are used to calculate the

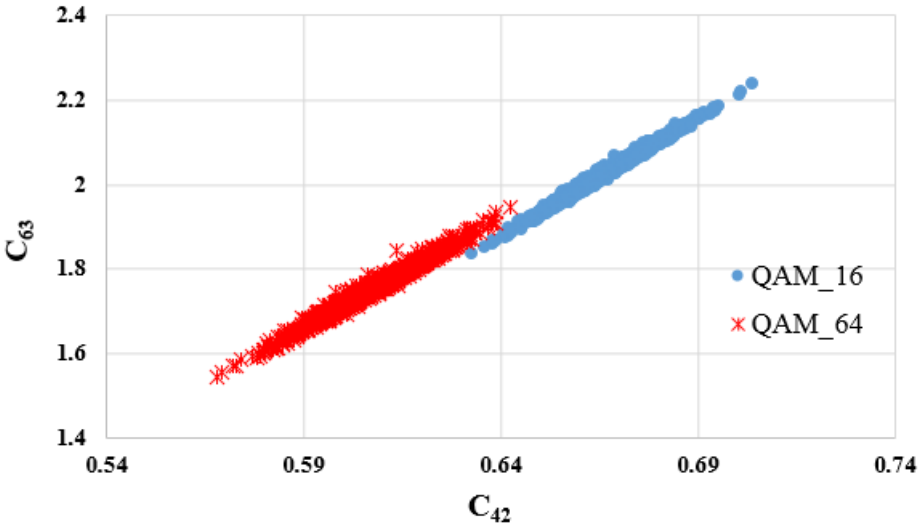


Figure 19: Features in the original dimensional space.

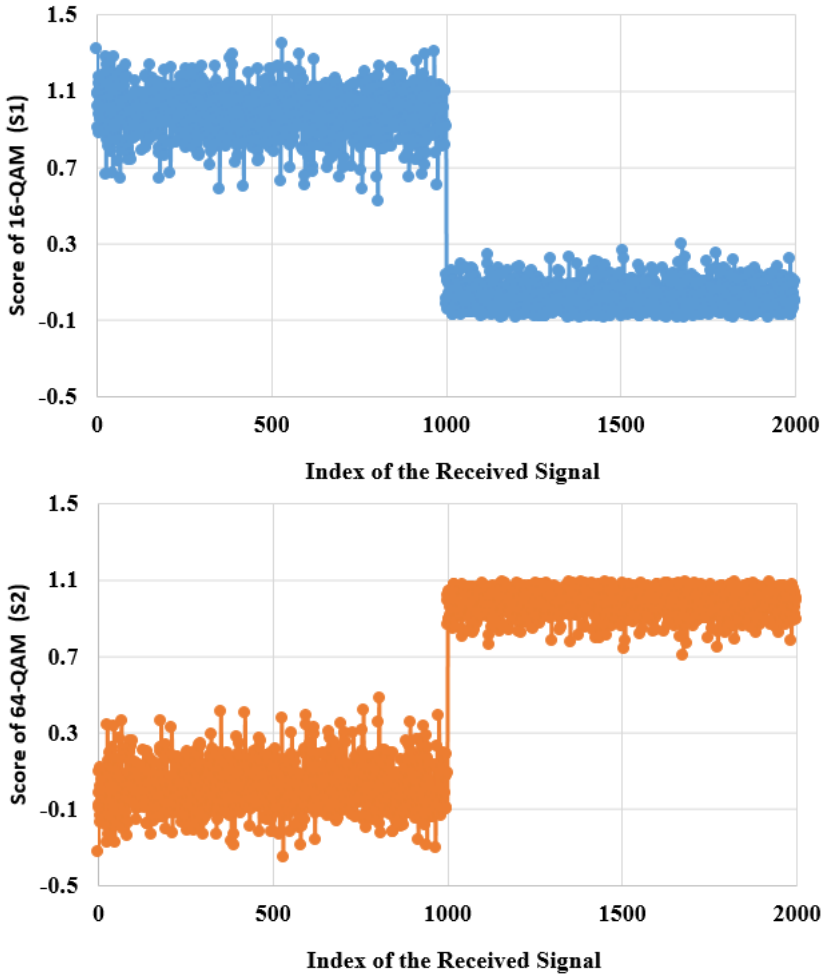


Figure 20: The two scores (super features) of the polynomial classifier.

average classifier weights. The average weights are then used during the testing stage to classify the modulation type of unknown signals regardless of the received signal SNR. The results show that the HP classifier provides better correct classification of the modulation type than the CP scheme, especially at higher values of SNR. However, due to the lack of SNR information, the performance of both schemes reaches an upper limit indicating the non-optimality of selecting the classifier weights. Then, we investigated the case when the SNR is known at the receiver. Accordingly, classifier weights are calculated for each SNR value in the training stage. Then, based on the SNR of the received signal, the corresponding classifier weights are used in the testing stage. The results presented in Figure 21 show that when the channel information is available at the receiver, a significant improvement in the classification accuracy is achieved, reaching to about 100% accuracy for SNRs above 12 dB. The advantage of using the proposed HP over the CP in the two scenarios is clearly demonstrated.

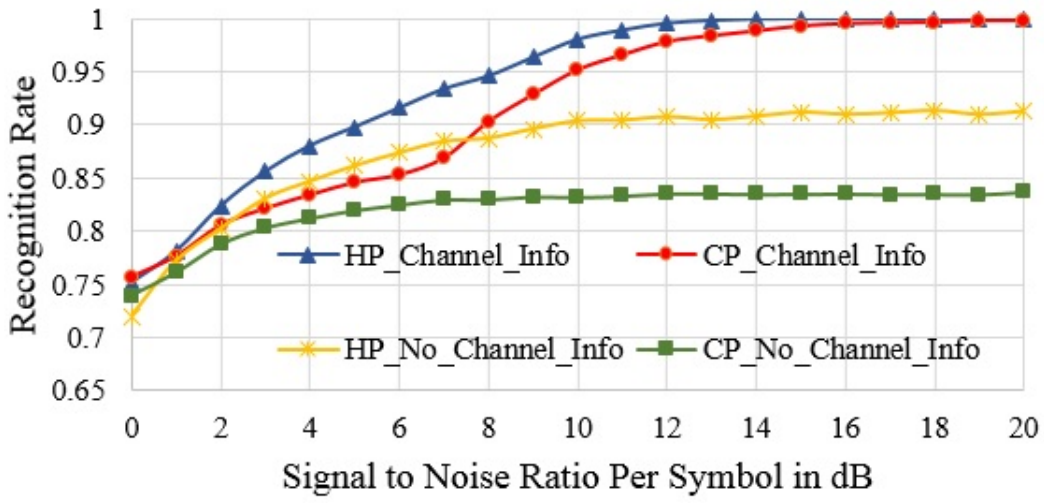


Figure 21: Classification rate in AWGN only for CP and HP classifiers.

The effect of using a different number of symbols per data block on the classification rate of an HP classifier is shown in Figure 22. It is shown that using a larger number of received symbols to extract the classification features improves the classification rate. For example, using N=5000 symbols leads to a 4 dB improvement in SNR compared to N=1000 symbols.

The performance of different classifiers is investigated for the same modulation

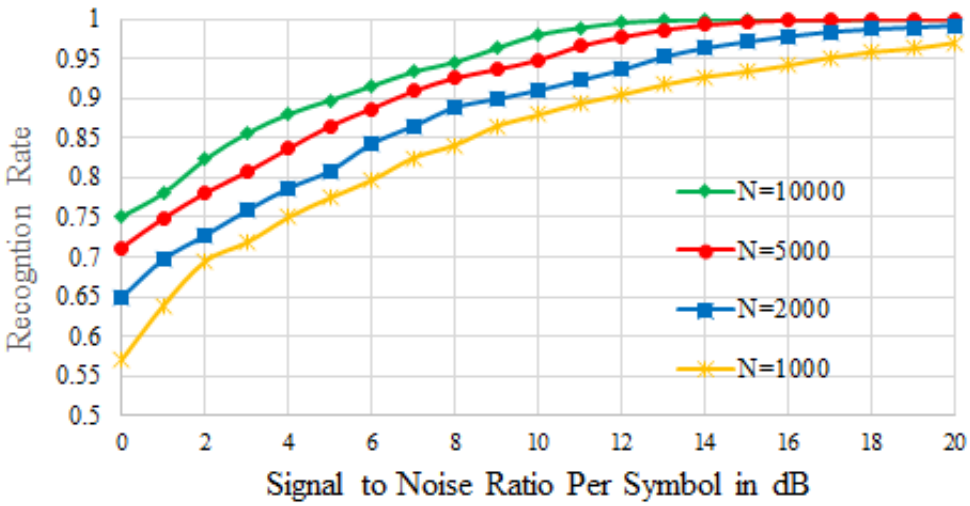


Figure 22: Classification rate in AWGN only for different numbers of symbols.

classification problem. In order to use the KNN classifier, the performance of the classifier is firstly tested for different values of K using different distances. This test is done using signals of length N=10000 at 10 dB. Figure 23 shows the classification accuracy when using Euclidian, city block, cosine and correlation distances. From Figure 23, it is obvious that using the correlation distance with K=16 results in the best classification accuracy.

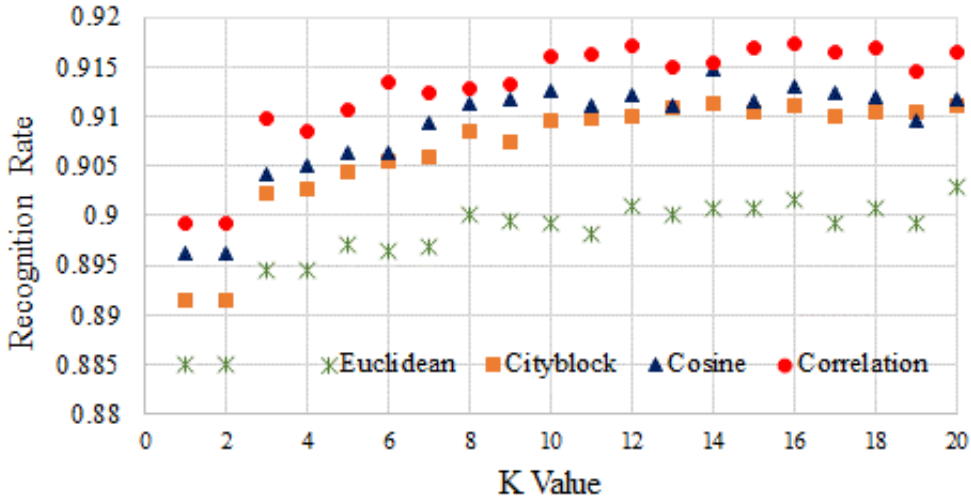


Figure 23: Classification using KNN classifier.

Next, the performance of the SVM classifier is examined for different values of

σ at 10 dB and for N=10000. As shown in Figure 24, the highest classification accuracy is achieved for σ= 2.40. Accordingly, a comparison between the different classifiers is conducted using SVM with σ= 2.40 and a KNN classifier with correlation distance and K=16.

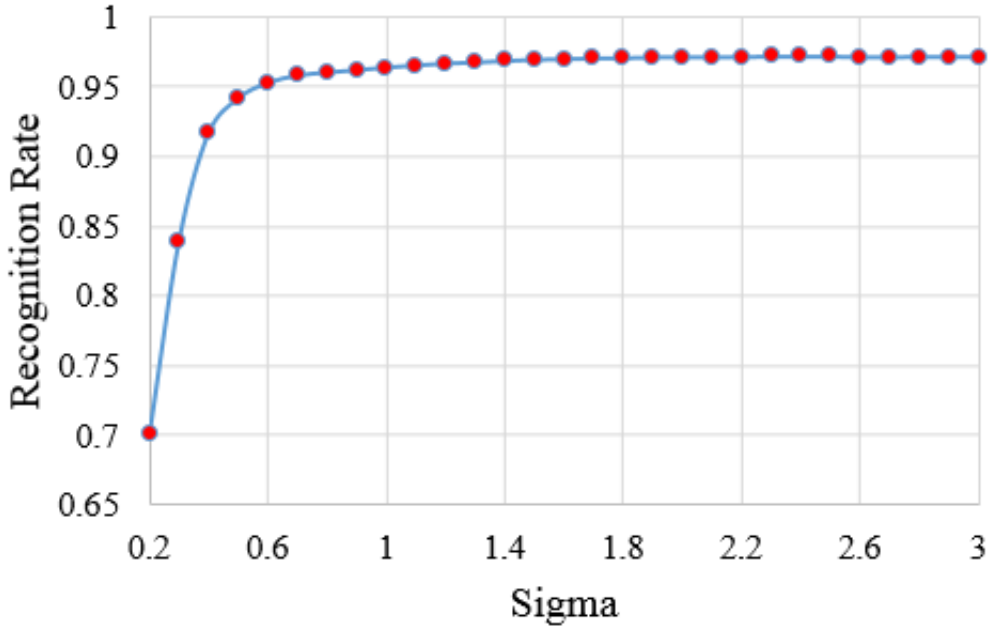


Figure 24: Classification using SVM classifier.

In Figure 25, the classification rate of the SVM, NN, naïve Bayes, KNN, and the proposed HP classifiers are presented and compared. The results are almost identical for SVM, NN and the proposed HP classifiers. The classification rate started to reach 100% at 13 dB. On the other hand, KNN and naïve classifiers achieved lower overall classification accuracy compared to the others, with the KNN classifier having more an advantage.

To summarize, the proposed classifier achieved the best accuracy in AWGN environments. Although NN and SVM have similar classification rates in this scenario, preferring one classifier over another depends on other factors like simplicity and ease of implementation.

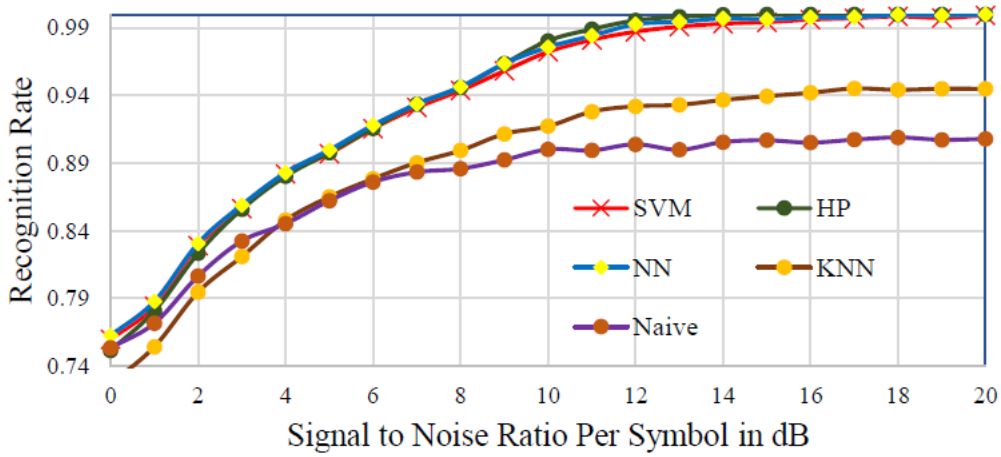


Figure 25: Comparison between different classifiers.

6.2. Classification in Slow Flat Fading Channels

In this section, the performance of the proposed hierarchical polynomial classifier is investigated in flat fading channel conditions with either Rayleigh (no line of sight) or Rician fading (with line of sight component). Without loss of generality, the Rician factor, which represents the ratio between the power of the direct path and the power of the reflected paths, is assumed to be 6. The parameters used are the same as in [20] with a sampling interval of 1×10^{-6} seconds, average path gains $[0, \quad -10]$ dB, path delays $\left[0, \quad 1 \times 10^{-7}\right]$ seconds, and maximum Doppler shift of 3Hz, 30Hz, or 83Hz. Figure 26 and Figure 27 show the recognition rate for N=5000 with different values of maximum Doppler shifts in Rician and Rayleigh fading, respectively. The maximum Doppler shift and the SNR of the channel are assumed to be known at the receiver side. Accordingly, these values are used to train the classifier. It is clear that the recognition rate is higher for Rician fading channels compared to Rayleigh fading channels. This is because of the existence of the strong signal due to the line of sight component. It is noted that a higher Doppler shift results in lower recognition rate.

The classification performance as a function of the number of received symbols and Doppler shift is shown in Figure 28 for flat Rayleigh fading. It is also observed that when the Doppler shift is high it is better to reduce the number of symbols used in the classification process in order to reduce the impact of the time variation of the channel on the classification performance. For instance, at a Doppler shift of 83 Hz,

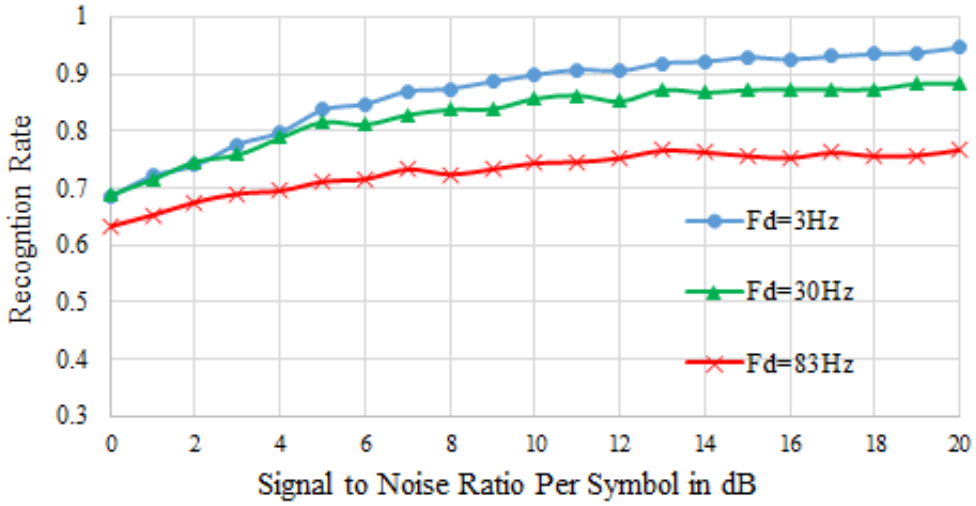


Figure 26: Classification rate in Rician fading channels.

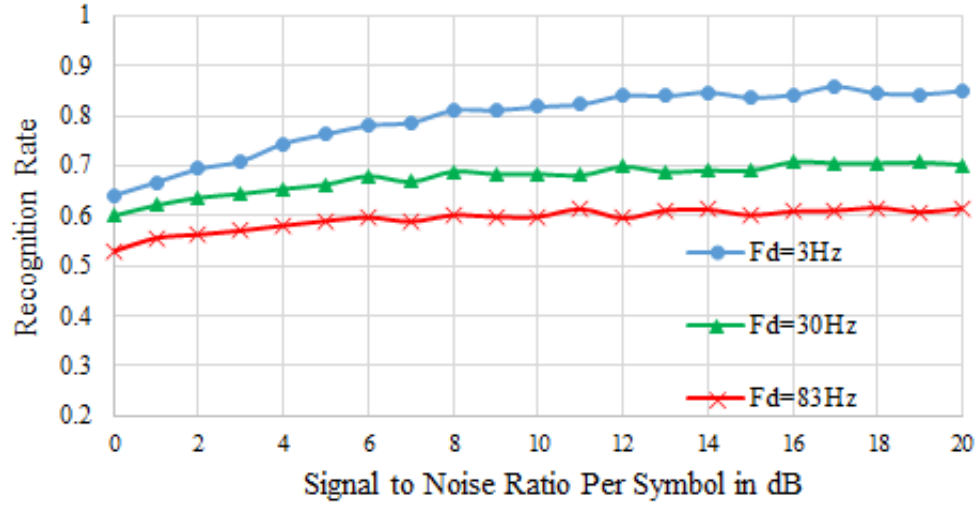


Figure 27: Classification rate in Rayleigh fading channels.

the classification rate reaches it is peak by using about 500 symbols as opposed to a Doppler shift of 30 Hz that requires about 1500 symbols to reach its maximum.

Figure 29 presents the constellation diagram for a QPSK signal plotted at different time instants. This explains the degradation in the performance when the number of symbols is increased for the case of high Doppler shifts. It is obvious that for N=20000 and N=10000, the constellations are very different compared to the original QPSK constellation; this is due to the fast variation in the channel during the time of receiving these symbols. However, for smaller numbers of received symbols, the constellation tends to reserve its main structure. Accordingly, to achieve high classification accuracy,

the number of received symbols should be optimized to be high enough to increase the classification accuracy and small enough to preserve the shape of the constellation. This optimized value of N varies based on the Doppler shift of the channel.

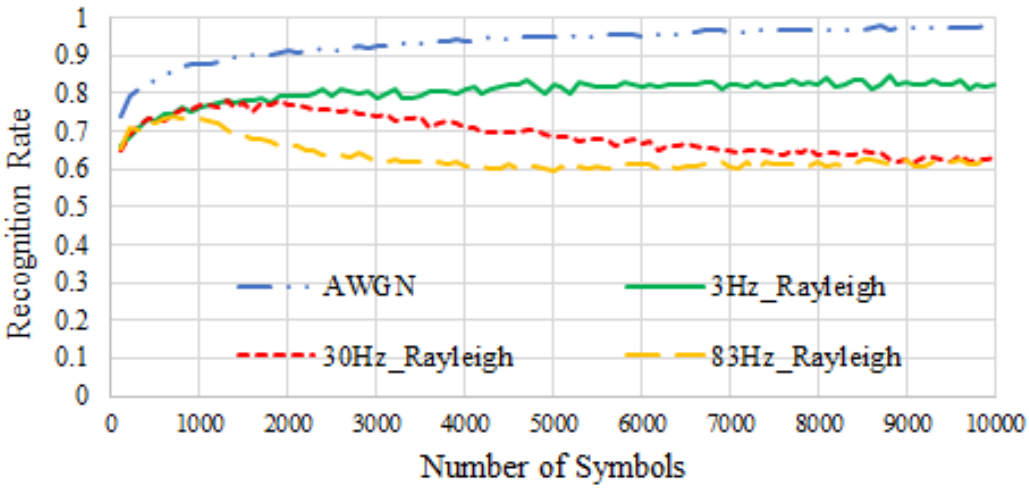
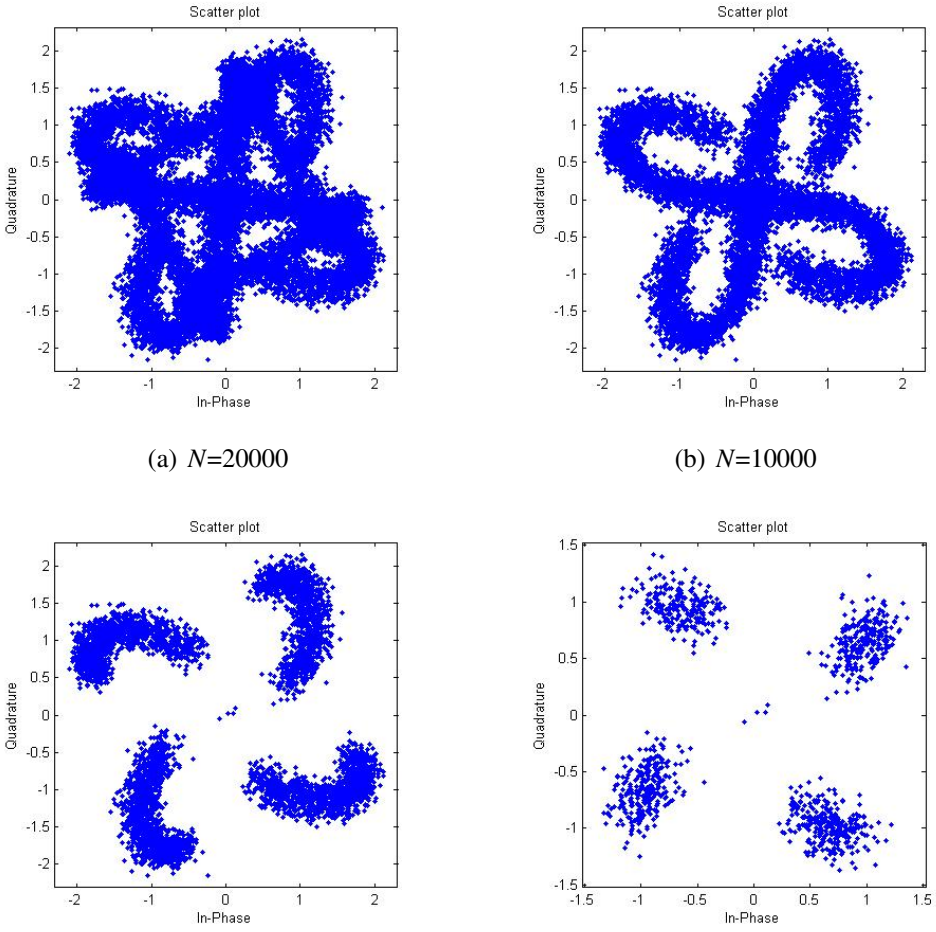


Figure 28: Classification rate versus number of symbols for different values of Doppler shift.

6.3. Classification Using Stepwise Regression

In this section, stepwise regression is integrated with the proposed hierarchical polynomial classifier with the aim of selecting the best classification features and reducing the system complexity. In the training stage, the second-order polynomial expansion of the input HOCs defined in Table 2 is calculated. The result is a new vector with fifty five classification features. These features are fed to the stepwise regression model in order to select the best features and discard the insignificant ones.

The values of α and β are chosen to be 0.01 for all the experiments. Figure 30 shows the recognition rate for a classification problem among BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, and 256-QAM. It is clear from the graph that the classification rate increases when the number of received symbols increases; for example, the correct classification accuracy is 97.94% for $N = 10000$ at 10dB compared to 91.08% and 87.93% at $N = 2000$ and $N = 1000$, respectively. Moreover, the recognition rate improved as



(a) N=20000 (b) N=10000

(c) N=5000 (d) N=1000

Figure 29: Constellation diagram for Rayleigh fading channel at 20 dB and for F_d=83Hz.

the SNR of the signal increases to reach 100% at 16 dB for $N = 10000$ and $N = 5000$.

Figure 31, Figure 32, Figure 33, Figure 34 and Figure 35 show the number of selected features at different SNR values for the five classification stages in Figure 16. In general, the number of selected features increases at high SNRs, since variation in the features' values is greater at low SNRs.

Moreover, the figures show the reduction in the number of classification features when using stepwise regression. For example, only three features are required to classify 64-QAM and 256-QAM in the range of 0dB to 3dB (see Figure 31), which indicates a 94.5% reduction in the number of features compared to the original scenario. On the other hand, in some cases the number of features reaches forty-four like

in Figure 31 at 20dB, where 80% of the original expanded feature vector are used. It is also worth mentioning that the stepwise regression produces some errors in selecting the final model, meaning that not all the significant features are selected, and not all the redundant features are discarded. This error is due to the inaccurate estimation of the coefficient values B_j during the regression process. Table 3 compares the classification rate obtained when using the original expanded feature vector as in Figure 22 to the case when the reduced set of features is used. Both systems provided the same classification accuracy, which confirms the possibility of having redundant or insignificant features that can be eliminated leading to reduced system complexity.

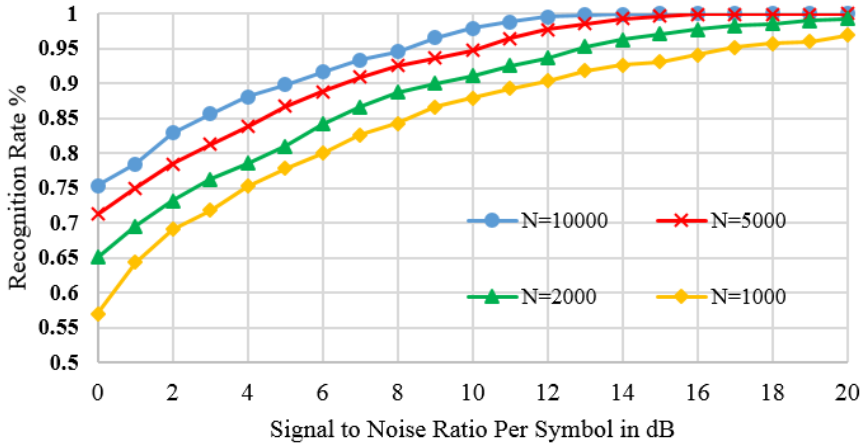


Figure 30: Classification Rate for Different Number of Symbols Using Stepwise Regression.

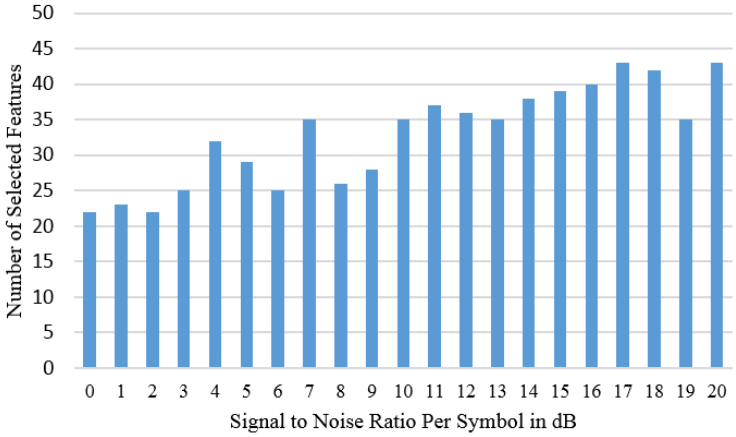


Figure 31: MPSK Versus MQAM.

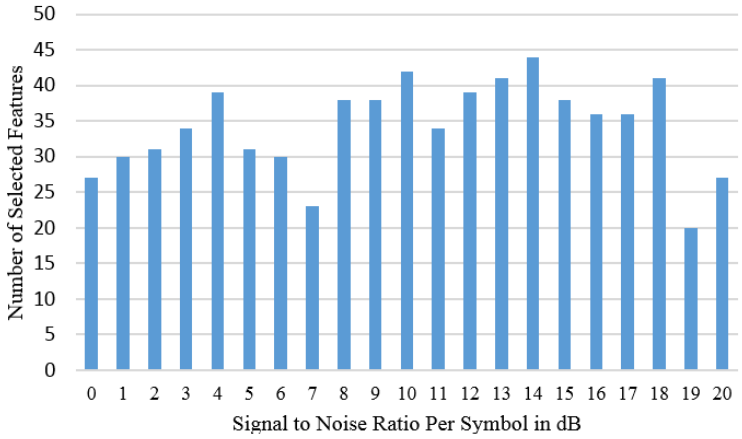


Figure 32: BPSK Versus QPSK and 8-PSK.

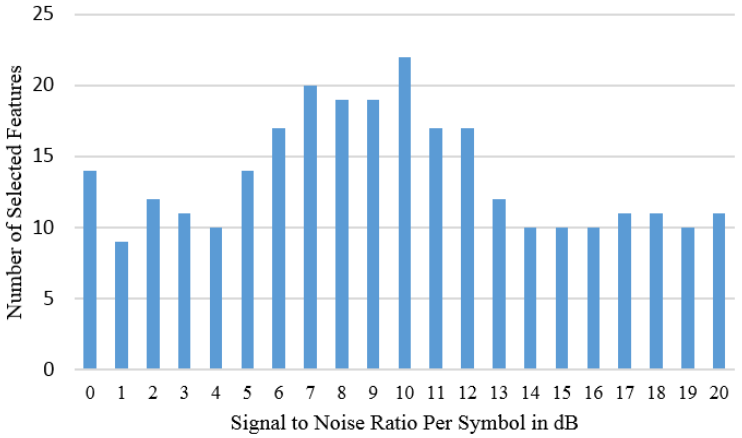


Figure 33: QPSK Versus 8-PSK.

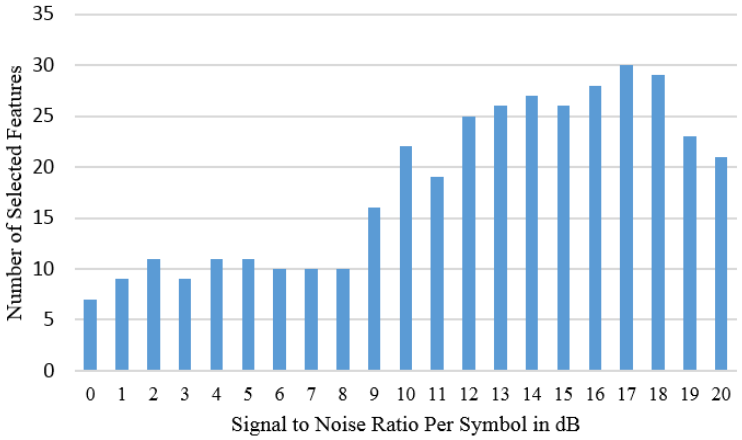


Figure 34: 16-QAM Versus 64-QAM and 256-QAM.

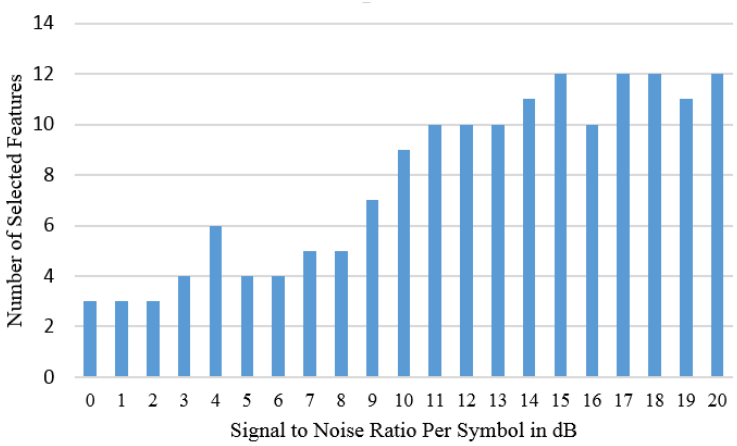


Figure 35: 64-QAM Versus 256-QAM.

Table 3: Stepwise Regression System Versus Normal HP Classifier System

System type	SNR (dB)					Number of Symbols
	0	5	10	15	20	
Stepwise	75.35	89.84	97.94	100	100	$N = 10000$
HP	75.14	89.78	98.04	100	100	
Stepwise	71.25	86.68	94.78	99.62	100	$N = 5000$
HP	71.22	86.53	94.84	99.67	99.98	
Stepwise	65.05	81	91.08	97.07	99.23	$N = 2000$
HP	64.96	80.88	91.03	97.16	99.2	
Stepwise	56.98	77.84	87.93	93.06	96.92	$N = 1000$
HP	57.08	77.48	87.93	93.38	97.04	

6.4. Classification Using Estimated SNR

As mentioned earlier, integrating the SNR estimation system with the proposed HP classifier improves the classification accuracy. However, to understand how this enhancement occurs, different scenarios are considered. In the first scenario, the system was trained using signals with SNRs in the range of 0 dB to 20 dB to calculate the first set of weights. The system was then tested with signals in the specified SNR range and the performance is shown in Figure 36. As the results show, the classification accuracy was about 70% at 0 dB and increased as the SNR increased. As for the second scenario, the system was trained using signals with SNRs in the range of 5 dB to 25 dB, where

signals between 0 dB to 5 dB were ignored. The system was then tested for signals with SNRs in the range of 0 dB to 20 dB. The result shows a degraded performance for signals with SNRs less than 5 dB while the performance, compared to the first scenario, increased for higher SNRs as shown in Figure 36 and Figure 37. Likewise, for the third scenario, signals with SNRs between 0 dB to 10 dB were ignored, and the training was conducted using signals with SNRs in the range of 10 dB to 30 dB. The performance of the system for this scenario shows further degradation for low SNRs compared to the first two scenarios, while for higher SNRs the performance is enhanced as shown in Figure 36 and Figure 37. The improvement in the classification accuracy was due to the exclusion of low SNR signals (noisy signals); therefore the training was optimized for clean signals. The previous discussion showed the importance of optimizing the training procedure in order to achieve the highest possible classification accuracy.

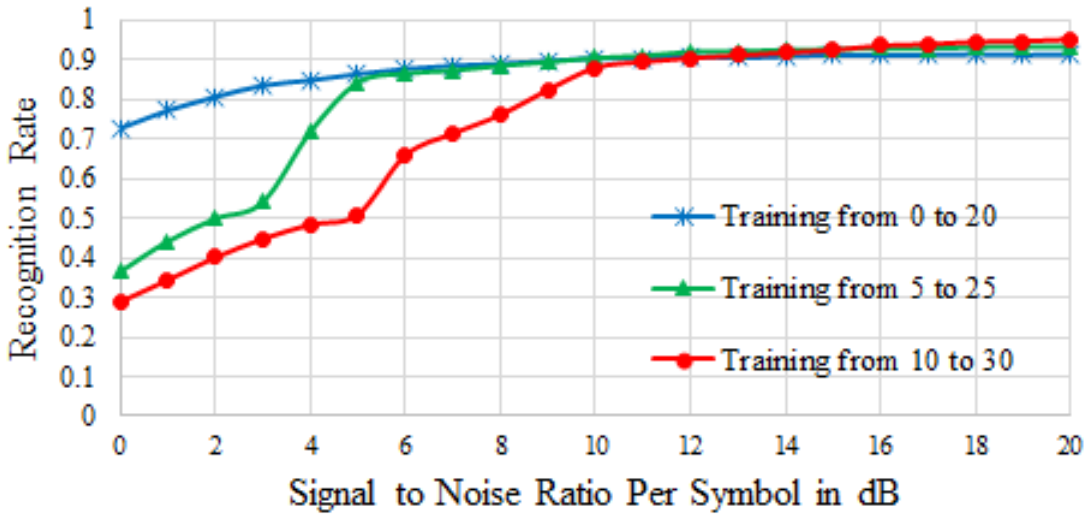


Figure 36: Training for different ranges of SNR.

Next, the performance of the proposed SNR estimator is examined by signals with SNR values in the range of 0 dB to 20 dB. Figure 38 shows a very accurate SNR estimation at low SNR values (approximately 100% correct estimation for SNRs in the range of 0 dB to 6 dB, whereas, for higher SNR values, the estimator started to produce small errors between the adjacent SNRs).

From Figure 37 and Figure 38, we concluded that training the classifier for

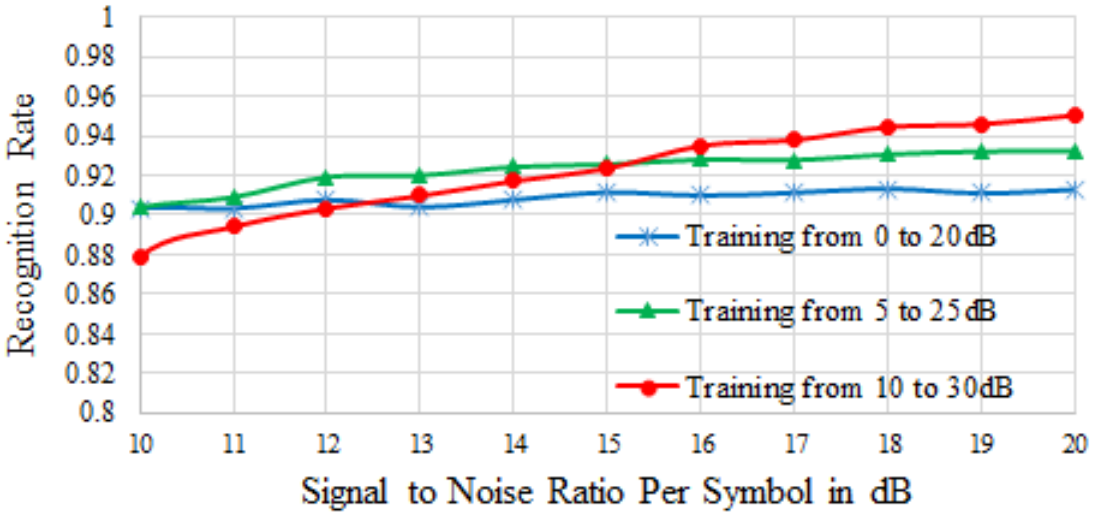


Figure 37: Training for different ranges of SNR (scaled).

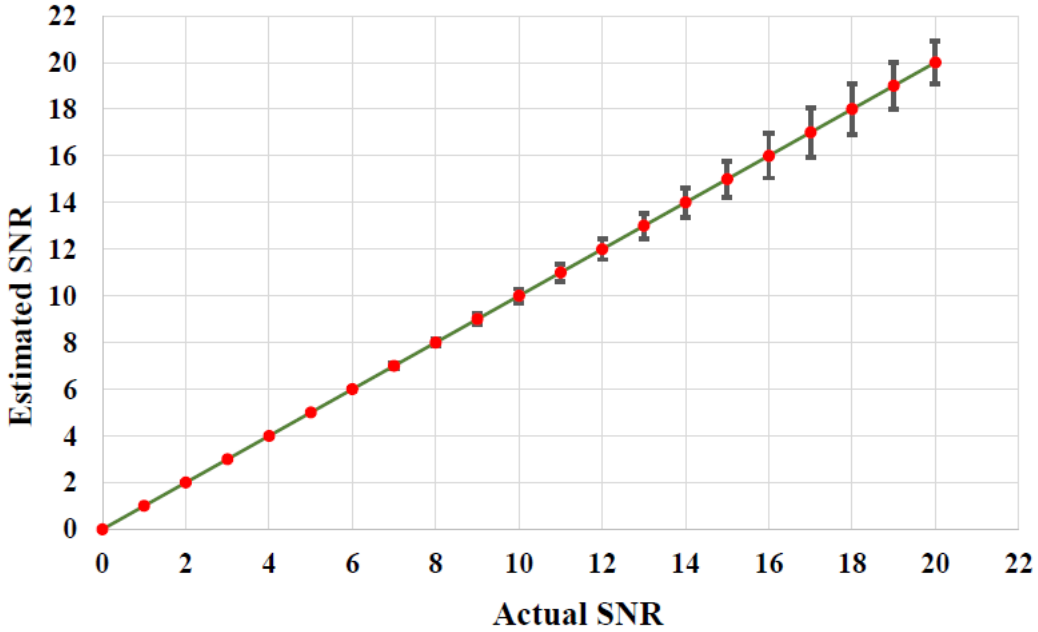


Figure 38: Estimated SNR versus actual SNR ($N = 10000$).

different ranges of SNRs changes the overall classification accuracy. Since the SNR estimator is relatively accurate for the SNR range of 0 to 10 dB, for this range, the estimated SNR is fed to the classifier and the corresponding automatic modulation classifier weights are used to determine the modulation type. If the SNR estimator identifies the SNR of the channel to be higher than 10 dB, then the average classifier weights for the range 10 to 30 dB are used. The result is improved classification accuracy as shown in Figure 39. It is clear from the graph that using a modulation classifier and SNR estima-

tor leads to the same performance (of the case of known SNR value at the receiver for low SNR values,) and results in an overall improvement of the classifier performance compared to the case of no channel information. Tables 4 to 6 show the confusion matrix of the SNR estimation system for the cases of $N = 20000$, $N = 10000$ and $N = 5000$, respectively.

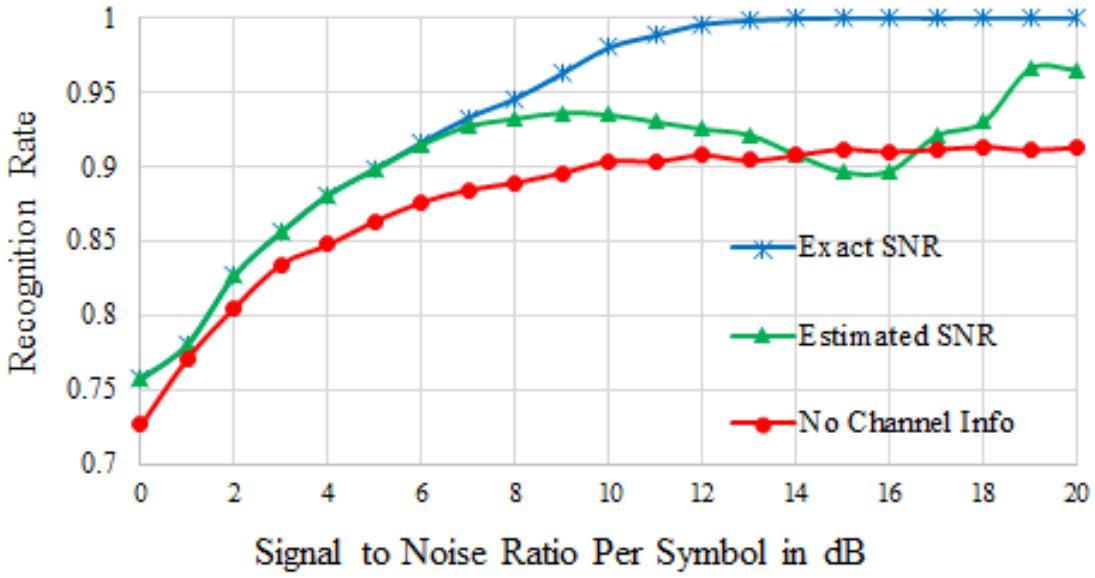


Figure 39: Classification using estimated SNR.

6.5. Comparison to Related Work

In this section, we compare the performance of the proposed system against other systems in the literature. Table 7 shows the classification rate of the Naïve Bayes classifier, Support Vector Machine (SVM) classifier, Maximum Likelihood (ML) classifier, Genetic programming with KNN classifier, and the proposed Hierarchical Polynomial (HP) classifier. To ensure a fair comparison, the proposed classifiers are simulated using the same parameters used by the other authors. In particular, we used four modulation types (BPSK, QPSK, 16-QAM and 64-QAM) with 512, 1024 and 2048 symbols per data block.

The proposed HP classifier achieved a classification rate of 100% at 20 dB for the three given scenarios, N=512, 1024, and 2048 symbols. For N=512 at 10 dB, the

Table 4: SNR estimation for $N = 20000$

	0 dB	1 dB	2 dB	3 dB	4 dB	5 dB	6 dB	7 dB	8 dB	9 dB	10 dB	11 dB	12 dB	13 dB	14 dB	15 dB	16 dB	17 dB	18 dB	19 dB	20 dB
0 dB	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 dB	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 dB	0	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 dB	0	0	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 dB	0	0	0	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 dB	0	0	0	0	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 dB	0	0	0	0	0	2	11998	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 dB	0	0	0	0	0	0	1	11992	7	0	0	0	0	0	0	0	0	0	0	0	0
8 dB	0	0	0	0	0	0	0	10	11964	26	0	0	0	0	0	0	0	0	0	0	0
9 dB	0	0	0	0	0	0	0	0	30	11900	70	0	0	0	0	0	0	0	0	0	0
10 dB	0	0	0	0	0	0	0	0	0	83	11717	200	0	0	0	0	0	0	0	0	0
11 dB	0	0	0	0	0	0	0	0	0	0	207	11406	386	1	0	0	0	0	0	0	0
12 dB	0	0	0	0	0	0	0	0	0	0	0	396	10961	638	5	0	0	0	0	0	0
13 dB	0	0	0	0	0	0	0	0	0	0	0	0	650	10408	927	15	0	0	0	0	0
14 dB	0	0	0	0	0	0	0	0	0	0	0	0	1	947	9689	1309	52	2	0	0	0
15 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1259	8987	1554	178	7	1	0
16 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84	1477	8324	1754	300	52	9
17 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	192	1654	7727	1796	477	153
18 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	372	1790	6966	1895	956
19 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	61	652	1793	6438	3054
20 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	222	748	1868	9146

Table 5: SNR estimation for $N = 10000$

	0 dB	1 dB	2 dB	3 dB	4 dB	5 dB	6 dB	7 dB	8 dB	9 dB	10 dB	11 dB	12 dB	13 dB	14 dB	15 dB	16 dB	17 dB	18 dB	19 dB	20 dB
0 dB	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 dB	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 dB	0	0	11999	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 dB	0	0	0	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 dB	0	0	0	0	11996	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 dB	0	0	0	0	3	11988	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 dB	0	0	0	0	0	14	11955	31	0	0	0	0	0	0	0	0	0	0	0	0	0
7 dB	0	0	0	0	0	0	35	11854	111	0	0	0	0	0	0	0	0	0	0	0	0
8 dB	0	0	0	0	0	0	0	84	11715	201	0	0	0	0	0	0	0	0	0	0	0
9 dB	0	0	0	0	0	0	0	0	205	11414	381	0	0	0	0	0	0	0	0	0	0
10 dB	0	0	0	0	0	0	0	0	0	394	10950	656	0	0	0	0	0	0	0	0	0
11 dB	0	0	0	0	0	0	0	0	0	0	595	10373	1023	9	0	0	0	0	0	0	0
12 dB	0	0	0	0	0	0	0	0	0	0	1	912	9695	1352	40	0	0	0	0	0	0
13 dB	0	0	0	0	0	0	0	0	0	0	0	12	1242	8966	1640	129	11	0	0	0	0
14 dB	0	0	0	0	0	0	0	0	0	0	0	0	43	1503	8397	1770	243	39	5	0	0
15 dB	0	0	0	0	0	0	0	0	0	0	0	0	1	145	1769	7607	1942	406	104	18	8
16 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	6	316	1918	6839	2102	504	199	116
17 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	1	57	515	2011	6134	2175	554	553
18 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	135	707	2011	5473	2098	1571
19 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	42	273	810	2065	4946	3863
20 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	143	492	893	2021	8441

Table 6: SNR estimation for $N = 5000$

	0 dB	1 dB	2 dB	3 dB	4 dB	5 dB	6 dB	7 dB	8 dB	9 dB	10 dB	11 dB	12 dB	13 dB	14 dB	15 dB	16 dB	17 dB	18 dB	19 dB	20 dB
0 dB	12000	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1 dB	1	11993 6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 dB	0	1	11987	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 dB	0	0	7	11969	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 dB	0	0	0	29	11919	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 dB	0	0	0	0	73	11767	160	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6 dB	0	0	0	0	0	140	11601	259	0	0	0	0	0	0	0	0	0	0	0	0	0
7 dB	0	0	0	0	0	0	250	11305	445	0	0	0	0	0	0	0	0	0	0	0	0
8 dB	0	0	0	0	0	0	0	417	10879	703	1	0	0	0	0	0	0	0	0	0	0
9 dB	0	0	0	0	0	0	0	0	663	10306	1020	11	0	0	0	0	0	0	0	0	0
10 dB	0	0	0	0	0	0	0	0	0	1002	9561	1392	45	0	0	0	0	0	0	0	0
11 dB	0	0	0	0	0	0	0	0	0	6	1233	8929	1719	104	9	0	0	0	0	0	0
12 dB	0	0	0	0	0	0	0	0	0	0	17	1542	8196	2013	207	22	2	1	0	0	0
13 dB	0	0	0	0	0	0	0	0	0	0	2	125	1865	7311	2221	393	66	15	2	0	0
14 dB	0	0	0	0	0	0	0	0	0	0	0	2	215	2070	6583	2401	494	177	36	15	7
15 dB	0	0	0	0	0	0	0	0	0	0	0	0	20	436	2131	6008	2301	638	237	123	106
16 dB	0	0	0	0	0	0	0	0	0	0	0	0	4	97	643	2276	5150	2372	657	359	442
17 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	13	228	794	2200	4714	2127	862	1062
18 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	3	85	403	933	2098	3819	2578	2081
19 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	184	526	849	1852	4116	4452
20 dB	0	0	0	0	0	0	0	0	0	0	0	0	0	1	15	102	309	655	759	2287	7872

HP classifier managed to achieve a 96.49% accuracy rate, whereas GP-KNN, SVM, and Naïve Bayes classifiers achieved an accuracy of 94%, 91.23% and 90.17%, respectively. On the other hand, only the ML classifier achieved a 75% classification rate. It is clear that the proposed HP classifier outperformed the four other classifiers in terms of classification rate. However, the classification rate is not the only factor that favors one classifier over another. A crucial factor is the system complexity and the order of calculations involved in making the classifier decision. Since modulation classification is mostly a real-time identification problem, systems with a simple structure and low complexity are preferred.

Table 7: Comparison to other systems in the literature

N	SNR (dB)	Classification Rate (%)				
		Naïve [44]	SVM [44]	ML [44]	GP-KNN [45]	HP (proposed method)
512	0	63.91	64.53	50	65	65.78
	10	90.17	91.23	75	94	96.49
	20	94.66	98.33	100	98	100
1024	0	69.68	70.3	50	70	71.99
	10	94.4	94.81	75	98	99.10
	20	98.28	98.89	100	100	100
2048	0	76.75	75.73	50	75	77.60
	10	97.89	97.92	75	100	99.96
	20	99.68	99.78	100	100	100

The complexity of a classifier can be analyzed based on the complexity of the training stage (offline) and the complexity of the testing stage (online). In the training stage, GP-KNN and SVM use iterative approaches and consume significant processing time to provide the final training model, especially for the case of GP-KNN where different function pools are used to calculate the final training model. However, for our proposed system and the naïve Bayes classifier, the training process is straightforward without any iterations involved. In general, the complexity of the training stage does not usually matter if the calculations in the testing stage are simple, because the training is performed offline.

In the testing stage, nine features are applied to the proposed HP classifier, and the classifier expands them forming a new feature vector with fifty-five features. It is noted that the only complexity in this method is in producing the expanded feature vector and then calculating the classifier scores. The complexity of the proposed HP classifier can be estimated as $O(2f)$ where f is the number of expanded features and the factor 2 is for the two binary sub-classifications required for each decision. However, the complexity of the proposed HP classifier can be reduced by using stepwise regression for feature selection. For the same classification problem, the combined system of the proposed HP classifier and stepwise regression has a complexity of $O(F_1 + F_2)$ where F_1 is the number of significant features for the first classifier to determine if the signal is MPSK or MQAM, and F_2 is the number of features for the second classifier to identify the exact identity of the modulated signal. For example at 20 dB, $F_1 = 33$ and $F_2 = 20$ for BPSK and QPSK, respectively, which indicates that the complexity of the system dropped significantly compared to the case without the feature selection model. It is worth mentioning that using stepwise regression adds to the system complexity in the training stage.

On the other hand, for the case of the GP-KNN system, the complexity is much higher since it involves the complexity of finding the final super feature (the calculations to find this super feature depend on the pool function used like plus, minus, times, reciprocal, negator, abs, sqrt, sin, cos, tan, asin, acos, or tanh [45]. Furthermore, in a GP-KNN classifier, the distance is calculated between the calculated super feature and all the reference samples. For example, the number of reference points in [14] is 300 points. Accordingly, after finding the super feature, the distances between this super feature and the 300 reference points are calculated. Hence, the overall complexity of GP-KNN is $O(N)$ where N=300 plus the complexity of generating the super feature.

For the SVM and naïve Bayes classifiers in the testing stage, the complexity is less than the GP-KNN classifier, but their overall classification accuracy is relatively lower. It should be noted that it is expected that the more complex computations the classifier performs, the higher classification rate it provides. Yet in our proposed system, we managed to achieve high classification accuracy and maintain simple classifier structure.

Chapter 7: Conclusion

Automatic modulation classification has a great importance in optimizing the usage of the available spectrum and increasing the data throughput of communication systems. Moreover, it has many other applications in military and civilian areas. In this thesis, an automatic modulation classification system using a hierarchical polynomial classifier and high-order cumulants as features is proposed. The proposed system is used to classify M-PSK and M-QAM modulations using binary sub-classification stages, where in each stage a separate polynomial classifier expands the original feature vector into a higher dimensional space in which the two considered classes are linearly separated.

A feature selection system based on stepwise regression is integrated with the proposed classifier, whereas the significant features for each binary classification stage are identified and used, and the insignificant features are removed. The result is a simplified modulation classification system with much reduced calculation complexity. Moreover, an SNR estimation scheme is introduced, where the SNR of the received signal is estimated and used to select optimized classifier weights. Integrating the SNR estimator with the proposed classifier results in an overall improvement in the classification accuracy of the system.

Different classifiers are simulated in this work, and the results showed that neural networks and support vector machines have very close performance to the proposed classifier. However, the proposed system is less complex, which makes it more suitable for real-time applications. For channels with slow flat fading, the proposed system is shown to have no degradation due to the phase shift in the constellation. Then, a relationship between the number of received symbols used to extract the classification features and the classification rate is investigated for different scenarios of Doppler shifts.

Finally, the advantage of the proposed system is investigated in terms of accuracy and calculation complexity compared to other work in the literature.

References

[1] H. Hansson, *ARTESA Network for Real-time Research and Graduate Education in Sweden*. Uppsala University, February 2006.

[2] B. Lathi, *Modern Digital and Analog Communication Systems*. Oxford University Press, 1998.

[3] A. Svensson, "An introduction to adaptive QAM modulation schemes for known and predicted channels," *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2322–2336, December 2007.

[4] B. Sklar, *Rayleigh Fading Channels*. CRC Press LLC, 1999.

[5] B. Sklar, "Rayleigh fading channels in mobile digital communication systems partI: Characterization," *IEEE Communications Magazine*, vol. 35, no. 7, pp. 90–100, July 1997.

[6] F. Belloni, "Fading models," S-88 Signal Processing Laboratory, HUT, Tech. Rep., 2004.

[7] H. Bertoni, *Radio Propagation for Modern Wireless Systems*. Prentice Hall Professional Technical Reference, 1999.

[8] S. Haykin and M. Moher, *Modern Wireless Communications*. Pearson Prentice Hall, 2004.

[9] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification using polynomial classifiers," in *25th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, September 2014, pp. 785–789.

[10] "Report of the spectrum efficiency working group," Federal Communications Commission Spectrum Policy Task Force, ET Docket 02-135.

[11] B. Ramkumar, "Automatic modulation classification for cognitive radios using cyclic feature detection," *IEEE Circuits and Systems Magazine*, vol. 9, no. 2, pp. 27–45, June 2009.

[12] M. Wong and A. Nandi, "Efficacies of selected blind modulation type detection methods for adaptive OFDM systems," in *1st International Conference on Signal Processing and Communication Systems (ICSPCS)*, December 2007, pp. 243–246.

[13] L. Häring, Y. Chen, and A. Czylwik, "Automatic modulation classification methods for wireless OFDM systems in TDD mode," *IEEE Transactions on Communications*, vol. 58, no. 9, pp. 2480–2485, September 2010.

[14] M. Aslam, "Pattern recognition using Genetic Programming for classification of diabetes and modulation data," Ph.D. dissertation, the University of Liverpool, Liverpool, 2013.

[15] B. Ramkumar, "Automatic modulation classification and blind equalization for cognitive radios ," Ph.D. dissertation, Virginia Polytechnic Institute and State University,Virginia, 2011.

[16] R. Li, "Modulation classification and parameter estimation in wireless networks," Master's thesis, Syracuse University, New York, 2012.

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

[18] C. Yin, B. Li, and Y. Li, "Modulation classification of MQAM signals from their constellation using clustering," in *10th Second International Conference on Communication Software and Networks (ICCSN)*, February 2010, pp. 303–306.

[19] A. Kubankova and D. Kubanek, "Algorithms of digital modulation classification and their verification," *WSEAS Transactions on Communications*, vol. 9, no. 9, pp. 563–572, September 2010.

[20] N. Geisinger, "Classification of digital modulation schemes using linear and nonlinear classifiers," Master's thesis, Naval Postgraduate School, California, 2010.

[21] M. Harbaji, "Classification of common partial discharge types in oil-paper insulation using acoustic signals," Master's thesis, American University of Sharjah, Sharjah, 2014.

[22] A. Swedan, "Acoustic detection of partial discharge using signal processing and pattern recognition techniques," Master's thesis, American University of Sharjah, Sharjah, 2010.

[23] W. Dan, G. Xuemai, and G. Qing, "A new scheme of automatic modulation classification using wavelet and WSVM," in *2nd International Conference on Mobile Technology, Applications and Systems, 2005*, November 2005, pp. 1–5.

[24] C. Yin, B. Li, and Y. Li, "Modulation classification of MQAM signals from their constellation using clustering," in *Second International Conference on Communication Software and Networks, ICCSN '10*, February 2010, pp. 303–306.

[25] A. Swami and B. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Transactions on Communications*, vol. 48, no. 3, pp. 416–429, March 2000.

[26] M. Mirarab and M. Sobhani, "Robust modulation classification for PSK /QAM/ASK using higher-order cumulants," in *6th International Conference on Information, Communications Signal Processing*, December 2007, pp. 1–4.

[27] N. An, B. Li, and M. Huang, "Modulation classification of higher order MQAM signals using mixed-order moments and fisher criterion," in *The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, vol. 3, February 2010, pp. 150–153.

[28] M. Aslam, Z. Zhu, and A. Nandi, "Automatic digital modulation classification using Genetic Programming with K-Nearest Neighbor," in *Military Communications Conference(MILCOM)*, October 2010, pp. 1731–1736.

[29] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," 1995.

[30] O. Konash and M. El-Sharakawi, "Economic dispatch using Particle Swarm Optimization for combined cycle generators," in *IEEE/PES Power Systems Conference and Exposition, PSCE '09*, March 2009, pp. 1–9.

[31] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. Academic Press, 2008.

[32] L. Zheng and X. He, "Classification techniques in pattern recognition," in *13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2005.

[33] N. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, August 1992.

[34] L. Ferrari, "Mining housekeeping genes with a Naïve Bayes classifier," Master's thesis, University of Edinburgh, Edinburgh, 2005.

[35] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, September 1943.

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

[37] P. Borwein and T. Erdelyi, *Polynomials and Polynomial Inequalities*. Springer, 1995.

[38] W. Campbell, K. Assaleh, and C. Broun, "Speaker recognition with polynomial classifiers," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 4, pp. 205–212, May 2002.

[39] Y. Hassan, M. El-Tarhuni, and K. Assaleh, "Comparison of linear and polynomial classifiers for co-operative cognitive radio networks," in *IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, September 2010, pp. 797–802.

[40] K. Assaleh and H. Al-Nashash, "A novel technique for the extraction of fetal ECG using polynomial networks," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 6, pp. 1148–1152, June 2005.

[41] T. Shanableh and K. Assaleh, "Feature modeling using polynomial classifiers and stepwise regression," *Neurocomputing*, vol. 73, no. 10-12, pp. 1752–1759, June 2010.

[42] R. Ghunem, K. Assaleh, and A. El-Hag, "Artificial neural networks with stepwise regression for predicting transformer oil furan content," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 19, no. 2, pp. 414–420, April 2012.

[43] V. Orlic and M. Dukic, "Automatic modulation classification: sixth-order cumulant features as a solution for real-world challenges," in *20th Telecommunications Forum (TELFOR)*, November 2012, pp. 392–399.

[44] M. Wong, S. Ting, and A. Nandi, "Naïve Bayes classification of adaptive broadband wireless modulation schemes with higher order cumulants," in *2nd International Conference on Signal Processing and Communication Systems, ICSPCS*, December 2008, pp. 1–5.

[45] M. Aslam, Z. Zhu, and A. Nandi, "Automatic modulation classification using combination of Genetic Programming and KNN," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2742–2750, August 2012.

Appendix A: HOCs versus SNR

This section shows the average values of HOCs for the six modulation types: BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM and 256-QAM at different SNR values, the values are averaged for $N = 10000$ received symbols. Figures 40 to 48 show the discrimination power of HOCs for classifying the different modulation types. Moreover, Figure 41 shows the ability of C_{21} on identifying the SNR of the signal regarding its modulation type.

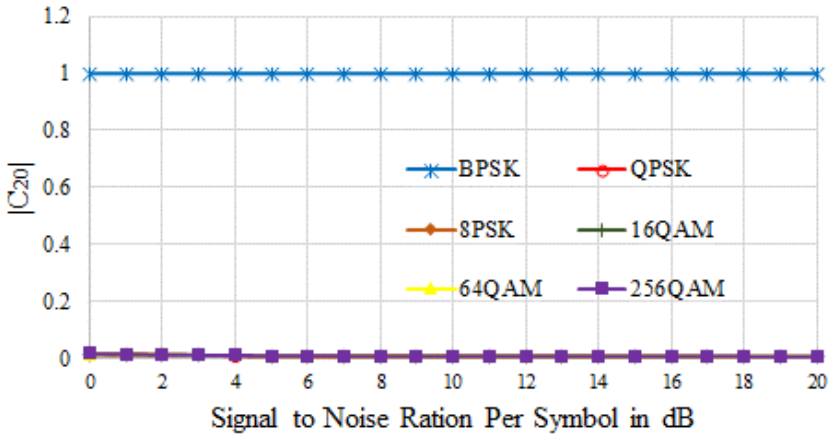


Figure 40: C_{20} versus SNR.

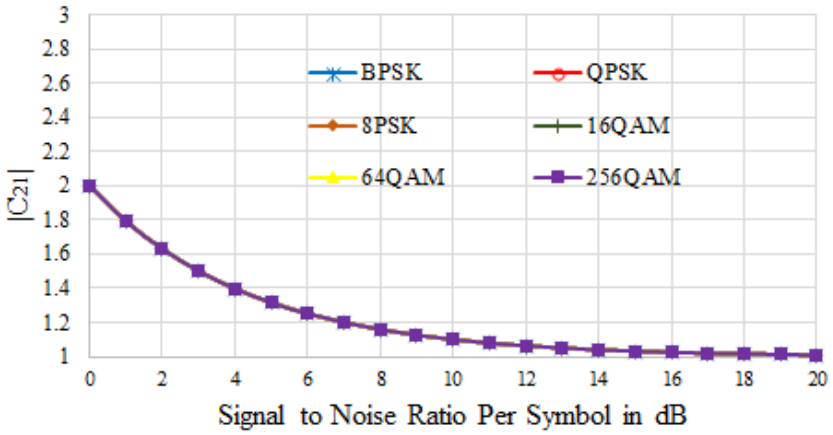


Figure 41: C_{21} versus SNR.

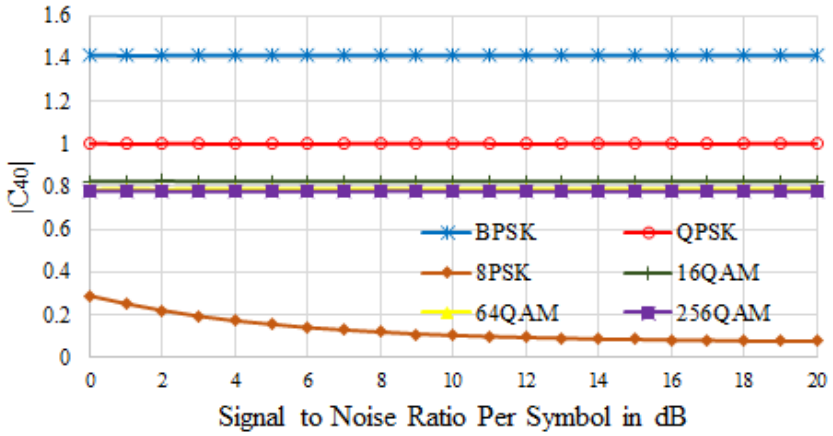


Figure 42: C_{40} versus SNR.

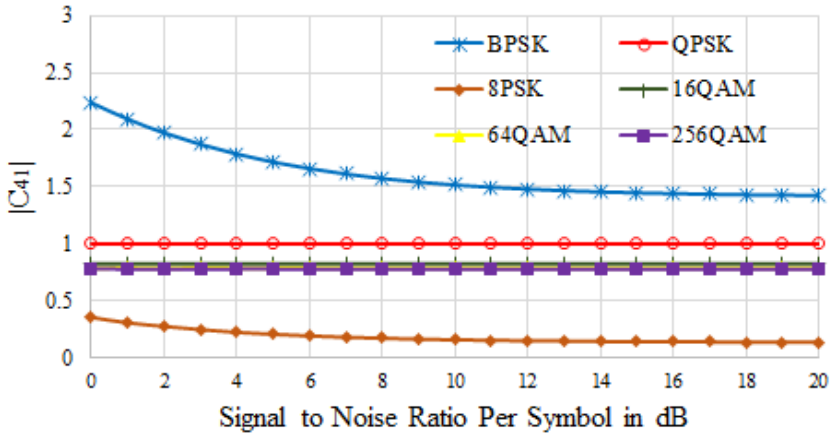


Figure 43: C_{41} versus SNR.

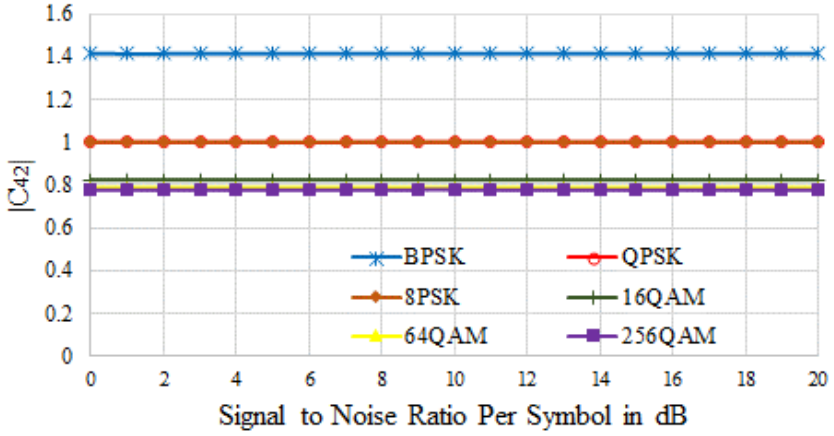


Figure 44: C_{42} versus SNR.

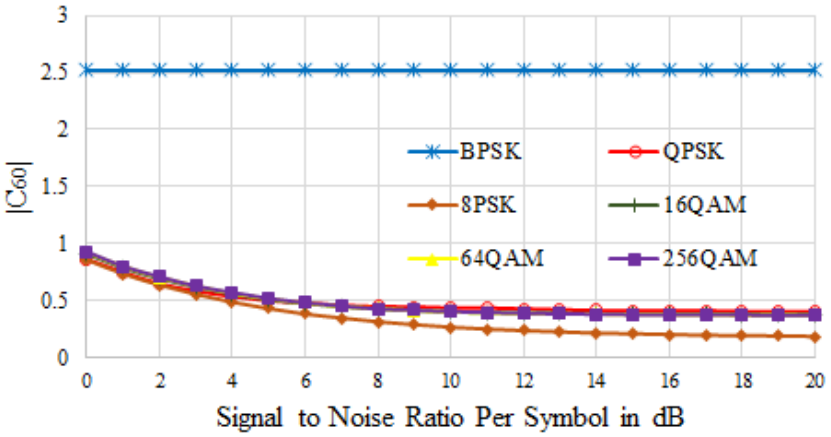


Figure 45: C_{60} versus SNR.

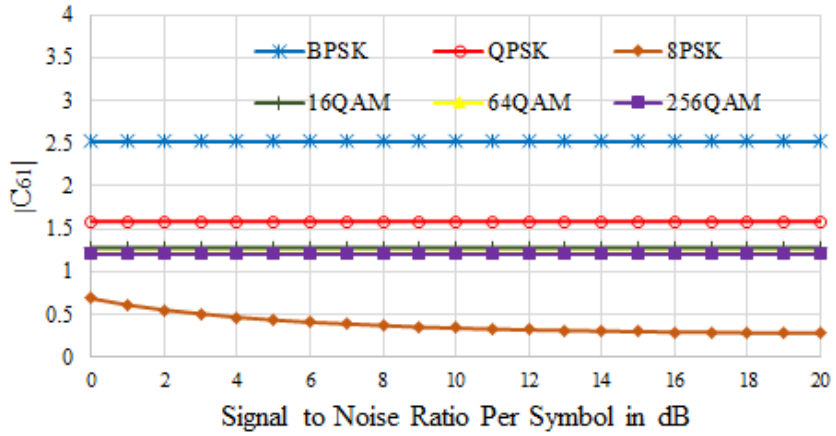


Figure 46: C_{61} versus SNR.

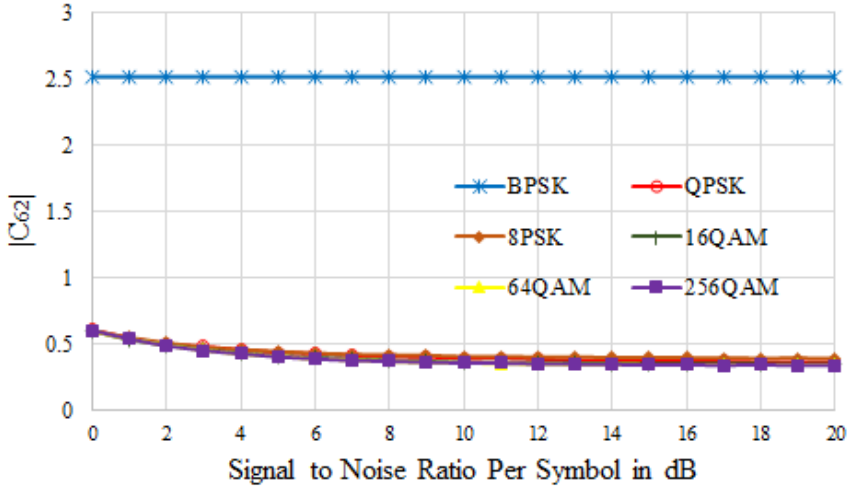


Figure 47: C_{62} versus SNR.

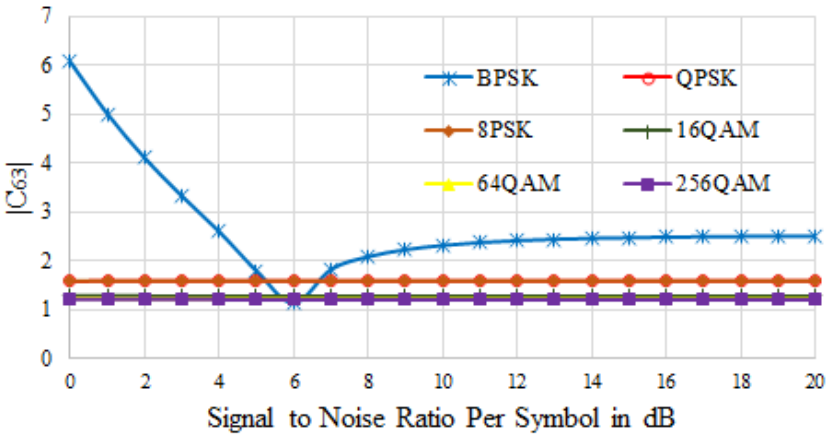


Figure 48: C_{63} versus SNR.

Vita

Ameen Abdelmutalab was born in 1988 in the United Arab Emirates, where he received his primary and elementary education. He then traveled to Sudan to study Electrical and Electronic Engineering at the University of Khartoum. In 2011 he received his Bachelor's degree and he started working as a teaching assistant in the Electrical engineering department at the University of Khartoum. In September 2012, Ameen started the Master's program at the American University of Sharjah. His research interests include wireless communications, cognitive radio, and machine learning applications.