

Article

Efficient Cumulant-Based Automatic Modulation Classification Using Machine Learning

Ben Dgani* and Israel Cohen* 

Andrew and Erna Viterbi Faculty of Electrical & Computer Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel

* Correspondence: bendgani@campus.technion.ac.il (B.D.); icohen@ee.technion.ac.il (I.C.)

Abstract: This paper introduces a new technique for automatic modulation classification (AMC) in cognitive radio (CR) networks. The method employs a straightforward classifier that utilizes high-order cumulant for training. It focuses on the statistical behavior of both analog modulation and digital schemes, which have received limited attention in previous works. Simulation results show that the proposed method performs well in different signal-to-noise ratios (SNRs) and channel conditions. The classifier's performance is superior to that of complex deep learning methods, making it suitable for deployment in CR networks' end units, especially in military and emergency services applications. The proposed method offers a cost-effective and high-quality solution for AMC that meets the strict demands of these critical applications.

Keywords: Cumulants; Modulation Classification; Machine Learning.

1. Introduction

The use of smart transmitters, adaptive receivers and intelligent communication networks has revolutionized radio communication in the modern era. These technologies have helped us optimize our usage of wireless channels. One crucial task in this regard is the implementation of intelligent radio modulation detection. Automatic modulation classification plays a vital role in various civilian and military applications, which are diverse and often massive. Established radio networks have been in operation for years and require fast and adaptive radio networks that can be used to combat various tasks daily and in times of need. To achieve this, the SDR platform is considered as a feasible option.

Software Defined Radio (SDR) is a communication platform that works through software programs rather than traditional hardware components. This means that the radio's functions can be adapted and changed dynamically based on different networks and tasks. SDR is considered the modern realization of the term "software radio," which was coined by Mitola [1]. By converting signal processing to the digital domain, SDR-based networks can incorporate complex algorithms that improve communication rate, accuracy, and distance.

One example of the benefits of SDR is the concept of Cognitive Radio, which was also introduced by Mitola in 1999 [2]. The idea behind Cognitive Radio is to develop a radio unit that can manage its communication network dynamically and automatically using software programs. The reason for this is to solve the problem of spectral congestion in modern communication. As the number of daily users increases, spectrum resources become more expensive and less available. By using Cognitive Radio, the radio unit can adapt to its environment and establish unlicensed communication networks that share the same wireless spectrum with other users. This helps solve the problem of spectral congestion and makes it easier for unlicensed users such as the military and emergency services to communicate effectively in times of crisis.

Citation: Dgani, B.; Cohen, I. Efficient Cumulant-Based Automatic Modulation Classification using Machine Learning. *Sensors* **2023**, *1*, 0. <https://doi.org/>

Received:

Revised:

Accepted:

Published:

Copyright: © 2024 by the authors. Submitted to *Sensors* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Setting up the shared network involves a method called spectrum sensing, where the Cognitive Radio unit can detect and analyze the available spectrum, and then manage the communication network accordingly. By utilizing software instead of expensive and complex hardware, Cognitive Radio offers a more efficient and cost-effective solution to the problem of spectral congestion. During the spectrum sensing step, a cognitive radio performs a scan of all available channels to find “empty holes,” or unused frequencies in the spectrum. By sensing channel parameters, the adaptive qualities of the cognitive radio enable it to dynamically change the modulation scheme used for transmission in order to improve transmission rates and quality. However, dynamic modulations require a recognition method for the received signal modulation scheme, which is known as automatic modulation classification, or modulation recognition and modulation detection in literature [3–5]. Spectrum sensing is essential for adaptive networks that must change according to their location without causing any interruption to local users’ daily communication needs, especially in times of emergency when every message counts.

Automatic Modulation Classification (AMC) is a process used to detect and identify the unknown modulation scheme of a radio signal. It has been around since the 1980s [6,7], but gained more popularity after the publication of the Cognitive Radio idea. Today, the modern AMC literature is mainly focused on deep learning and neural networks[8–10]. However, the classic AMC algorithms are split between two methods [12]. The process of identifying radio signals can be challenging due to the high computational complexity that certain algorithms require. The first set of methods, called likelihood-based algorithms [11], are accurate but can be expensive to implement, which is a drawback for government-funded services that rely on cost-efficient projects.

On the other hand, Feature-based (FB) methods offer a more efficient approach to signal analysis. By extracting features like high-order statistical moments (HOS) or instantaneous frequency and phase parameters, these algorithms can help understand signal characteristics. FB algorithms rely on a predetermined set of features, reducing the computational burden when compared to more complex machine learning techniques. To use FB algorithms, the process starts by extracting classification features from the signal. Once the algorithm has identified the key features, it can train the selected classification algorithm. Some examples of feature-based algorithms are support vector machines (SVM), decision trees, and K-means. After the model is trained, it can be applied with minimal complexity and runtime, making it a useful tool for identifying radio signals.

In recent years, the field of neural networks has experienced massive growth, leading to most new studies on AMC utilizing neural networks for feature extraction, classification, or both. Neural networks have demonstrated optimal results concerning classification accuracy, achieving high detection probability even at low SNR values. However, like likelihood-based methods, neural networks can suffer from high computational complexity and require significant system storage if a deep network model is used. This can become a challenge when dealing with a quick-setup radio network that relies on end units and needs to be power- and time-efficient. This is particularly important when considering emergency services where there is a need for a new, easy-to-set-up network that can quickly adapt itself to the spectral environment. Therefore, the training process for the desired modulation must be efficient.

This paper focuses on the FB method for the AMC, where we utilize high-order cumulants (HOCs) as key features. Cumulants are statistical measures that are alternatives to moments such as mean and variance. The HOCs are popular in AMC literature as a detection feature [14,15], due to their additive property that separates the desired signal from its additive noise. While most literature discusses digital modulation classification [16,17], our work aims to include analog modulations as well. In our study, we analyze the statistics of eleven modulation schemes, explore the best classification feature for each modulation, and discover the statistical connection between pairs of modulations.

The rest of this paper is organized as follows. Section 2 presents the modulated signal models and the theory behind HOS and cumulants. In Section 3, we evaluate the simulation

results for our proposed FB method. Finally, in Section 4, we discuss the conclusions from the results.

2. Background and Algorithm

This section is divided into four subsections to provide a comprehensive explanation of our work. The first part briefly reviews the theory of high-order statistics and the cumulants, which provide an alternative approach to the more familiar moments. In the second part, we introduce the dataset that we used for our work. The third subsection explains the theoretical aspect of the classification process. Finally, the last subsection briefly explains the method we used for our work.

2.1. High Order Statistics

In statistics, moments of a random variable provide quantitative measures related to the shape of its probability density graph and the distribution behavior. Specifically, the n th order moment of a random variable x is defined by

$$m_n(x) = \int x^n f(x) dx, \quad (1)$$

where $f(x)$ is the probability density function of x . If the n th order moment integral diverges, the n th order moment of the variable does not exist. If the n th order moment of x exists, so does the $(n - 1)$ th order moment, and thus all lower order moments. We define a central moment by $m_n(x - \mu)$, which is the mean value of x , and define the normalized moment as $m_n(x/\sigma)$ where σ is the standard deviation.

The cumulants c_n of a probability distribution are an alternative to the moments of the distribution. Any pair of probability distributions with identical moments will also share identical cumulants, and vice versa. The first, second, and third cumulants are equal to the first three central moments - known as the mean, variance, and skewness, but fourth and higher-order cumulants diverge from the central moments. The cumulant of a random variable x is defined using the cumulant-generating function

$$C_t(x) = \log(E[e^{tx}]), \quad (2)$$

and the cumulants c_n are obtained from a power series expansion of the cumulant generating function:

$$C_t(x) = \sum c_n \frac{t^n}{n!}. \quad (3)$$

This expansion is a Maclaurin series expansion of the logarithm function, so we get that the n th-order cumulant can be obtained by differentiating the expansion n times at $t=0$. The use of cumulants instead of moments as features for AMC in the literature is usually based on their cumulative property: assume X and Y are independent random variables, then

$$c_n(X + Y) = c_n(X) + c_n(Y). \quad (4)$$

Despite its benefits, the cumulant generating function can be challenging to calculate, especially when we are not familiar with the exact distribution of the signal. However,

since moments and cumulants are shared when the distribution is the same, we can use the moments to calculate the cumulants based on the work in [15,18]. We use

$$c_{4,0} = m_{4,0} - 3m_{2,0}^2 \quad (5)$$

$$c_{4,1} = m_{4,1} - 3m_{2,1}m_{2,0} \quad (6)$$

$$c_{4,2} = m_{4,2} - |m_{2,0}|^2 - 2m_{2,1}^2 \quad (7)$$

$$c_{6,0} = m_{6,0} - 15m_{4,0}m_{2,0} + 30m_{2,0}^3 \quad (8)$$

$$c_{6,3} = m_{6,3} - 9c_{4,2}m_{2,1} - 6m_{2,1}^2 \quad (9)$$

$$c_{8,0} = m_{8,0} - 28m_{6,0}m_{2,0} - 35m_{4,2}^2 + 420m_{4,0}m_{2,0}^2 - 630m_{2,0}^4 \quad (10)$$

where $m_{q,p}$ is the mixed moment of the signal, defined as $E(x^q * \bar{x}^{p-q})$. This set of equations enables simple and fast calculation for the HOC, which will act as features for the classification algorithm. Since our work is based on digital signals, we will use the discrete approximation for the mixed moment of N samples from a signal:

$$m_{q,p} = \frac{1}{N} \sum_{n=1}^N x^q[n] * \bar{x}^{p-q}[n]. \quad (11)$$

The classification method in our paper uses the cumulants above as input features, along with the matching cumulants $C_{p,q}(x_d)$ for the derivative of the signal

$$x_d[n] = x[n+1] - x[n]. \quad (12)$$

After experimenting with different features, we added the cumulants of the derivative to the features list, including separated cumulants for the real and imaginary parts of the signal and the cumulants of the amplitude and phase of each signal. Only the derivative's cumulants have noticeably improved the classification accuracy.

When discussing AWGN (Additive White Gaussian Noise), the primary benefit of utilizing cumulants becomes evident. The noise is a random process that follows a normal distribution. Therefore we can use its cumulant generating function, which is calculated in [19,20] using

$$C_t(w) = \log(M_t(w)) = \mu t + \sigma^2 \frac{t^2}{2}, \quad (13)$$

where μ and σ are the mean and standard deviation, respectively. The function in (13) is a second-order polynomial, so its 3rd and higher-order derivatives are zero-valued, meaning the HOC cumulants of AWGN are equal to zero. Hence, The HOC of the sum between AWGN and another independent random process will equal the HOC of the second random process alone. Since the derivative of AWGN is distributed the same, the derivative's cumulants are still effective for our method under the above assumptions.

2.2. Modulation and Signal Model

The signals used in this paper are taken from the RadioML2018.01A dataset [21]. This dataset contains 2,555,904 frames. Each frame is 1024 IQ samples from a modulated signal, with various modulation schemes and SNR values from -20 to 30 dB. The dataset contains 24 different modulations generated using SDR under the following channel specification:

- Selective multipath Rician fading;
- Carrier frequency offset;
- Symbol rate offset;
- Non-impulsive delay spread;
- Doppler shift;
- AWGN.

We only consider 11 modulations as the "Normal" set. This includes the following: OOK, 4ASK, BPSK, QPSK, 8PSK, 16QAM, AM-SSB-SC, AM-DSB-SC, FM, GMSK, OQPSK. (We

will refer to AM-SSB-SC and AM-DSB-SC as SSB and DSB, respectively). We decided to evaluate our work on the “Normal” dataset since it is considered the benchmark for classification models in the literature, in contrast to the “hard” set, which contains all 24 schemes.

2.3. Decision Tree

A decision tree is a flowchart-like model for supervised learning in which each internal node represents a decision test on a single feature, each branch represents the outcome of the test, and each leaf node represents a class label. The paths from root to leaf represent classification rules. Decision trees have several advantages as a classification model, and we are focusing on two of them in this paper. The first is that training the tree model does not require a large dataset; this is a significant advantage over neural networks that are widely popular in the literature despite the need for massive data to yield results. The second advantage is the model’s simplicity, which allows it to combine with other decision techniques. The benefit of combining techniques comes when we make a tree model, where instead of a simple decision rule at each node, we can use a more complex model.

The decision tree model we used in our work is from the Python scikit-learn (Sklearn) library, which is trained based on Gini impurity. Gini impurity is a measurement for the probability of classifying the labels incorrectly. Let A be a random variable that takes one of k labels (a_1, a_2, \dots, a_k) ; then we define the Gini impurity of a classification model as

$$G(A) = 1 - \sum_{i=1}^k (P(a_i))^2. \quad (14)$$

The training process of a decision tree contains three steps:

1. Decide feature to split the data: for each feature, the Gini impurity is calculated, and the one for which it is minimized is selected;
2. Continue splitting the data based on 1;
3. Stop splitting the data if we reach a certain tree depth.

The Decision tree model’s parameters are the comparison threshold at each node, barring the leaves, meaning the total number of parameters in a tree of depth d can be calculated with

$$P_n = 2^H - 1, \quad (15)$$

this is also the number of calculations a single tree applies when classifying since there is no need for multiplications or additions.

2.4. Method

Our classification algorithm is designed for integration within SDR systems operating in Cognitive Radio networks. The workflow initiates with signal acquisition by the SDR, followed by the segmentation of acquired signals into discrete packets, typically standardized at a length of 1024 samples — although this length is variable, subject to network requirements and design. These packets serve as inputs to our classification algorithm. For each packet received, we use the set of equations (5) through (10) to extract the features vector of the cumulants. This feature extraction process is computationally straightforward, involving a singular pass through the packet data, which aligns with the predetermined packet length of the communication network’s settings.

Subsequently, these extracted features undergo the classification phase, where the decision tree model is employed. The output of this classification stage yields the identification of one among the 11 available modulation schemes within the “Normal” modulations set. The detected modulation holds significant utility within the Cognitive Radio context, facilitating dynamic functionalities such as adaptive frequency channel selection and network usage detection. Furthermore, the system’s training procedure can be executed proactively in controlled environments or dynamically during operational use, enabling

real-time adaptation to the evolving network landscape. This adaptability ensures the system is efficient and adaptive within varying operational scenarios, allowing seamless integration into dynamic spectrum environments.

3. Simulation and Results

The suggested classification method requires hand-picked features for each step, so we start the simulation with a small-scale experiment. We start with an experiment to determine which of the features is optimal for the classification and detection of every single modulation from our desired set of 11 schemes. This simulation includes a simple decision tree model, where only one feature from the cumulants and derivative cumulants is used each run, where a run is defined as training and testing. Table 1 shows the top two features for each modulation scheme, in terms of detection accuracy using the simple decision tree model.

Table 1. Best two features for each modulation at SNR=10dB.

Modulation	1st Feature	P_d	2nd Feature	P_d
SSB	$C_{4,1}$	100%	$C_{4,0}$	100%
DSB	$C_{4,1}^d$	74.69%	$C_{4,0}^d$	73.34%
BPSK	$C_{2,0}$	99.57%	$C_{4,0}$	92.73%
FM	$C_{4,1}$	99.63%	$C_{2,0}$	99.27%
GMSK	$C_{4,2}$	96.09%	$C_{8,0}$	81.23%
OOK	$C_{2,0}$	95.82%	$C_{2,1}$	93.61%
4ASK	$C_{2,1}$	95.84%	$C_{2,0}$	93.71%
QPSK	$C_{4,0}^d$	87.95%	$C_{8,0}^d$	86.69%
OQPSK	$C_{4,2}$	76.39%	$C_{4,0}$	60.53%
16QAM	$C_{6,0}^d$	100%	$C_{6,3}^d$	99.74%
8PSK	$C_{4,0}^d$	100%	$C_{4,2}^d$	99.87%

Table 1 shows the average outcomes obtained when classifying signals specifically at an SNR of 10dB. This SNR value is deemed sufficient for a receiver to effectively detect and demodulate signals with accuracy. Utilizing the insights derived from this table, we proceed to construct the ultimate feature vector, which forms the foundational input for our classification algorithm.

This feature vector is meticulously composed of a consolidated set of cumulants, as illustrated in the table, representing a combined and optimized selection based on their performance at the specified SNR of 10dB. This selection process is pivotal in crafting an efficient and discriminative feature set, chosen explicitly for its effectiveness in signal classification within this specific SNR context. By amalgamating these cumulants into a unified feature vector, we aim to encapsulate the most informative and distinguishing signal characteristics essential for accurate modulation classification at this SNR threshold.

The next part of the simulation is the picking process of the training set for the model. It is widely agreed that for training machine learning models, the algorithm randomly picks the training set, out of the total database. However, in our work, we consider training our model with only a subset of the data. This subset includes all the sampled signals with SNR values higher than a certain threshold. The reasoning behind our decision comes when we study different modulation schemes, we can see that each scheme has a minimal SNR requirement for adequate demodulation, which means that using signals with SNR below that threshold might affect our model training in negative ways. Based on these facts, we split the data into two separate groups, high SNR signals and low SNR signals. In order to pick an SNR threshold for the splitting, we conducted our second experiment. Figure 1 shows the classification results when we train the model using a signal with SNR higher than varying thresholds, we also included a zoom-in view on the section where classification accuracy is the highest.

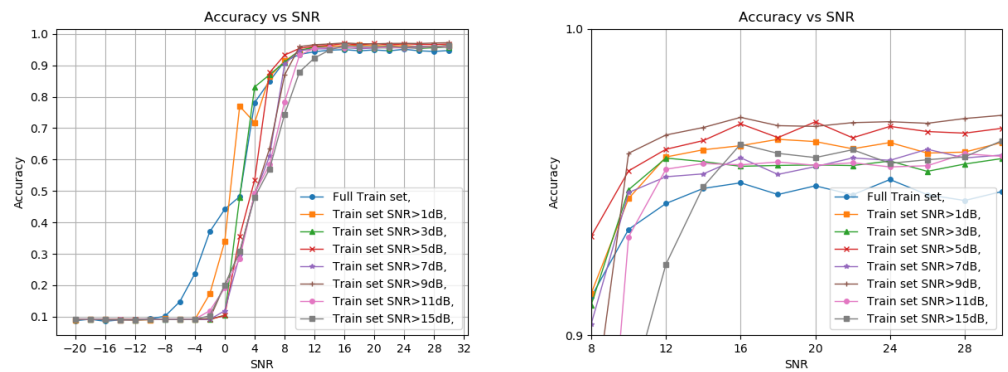


Figure 1. P_d based on train set SNR threshold. (a) Full view; (b) Zoomed in view.

First, picking a train set from all the SNR values gives better results at negative SNR but deteriorates when reaching a positive value, as expected. Another conclusion is that using over 5 dB as the threshold gives lesser accuracy for high and low SNR test samples. The final choice of threshold was between 1 dB and 5 dB. Although the 1 dB threshold yields better accuracy at lower SNR levels, a slight improvement in detection accuracy is observed starting from 4 dB and higher with the 5 dB threshold. This result supports the assumption that modulation schemes have minimal SNR requirements for perfect demodulation. The depth of the decision tree was not considered for the first test. So, we want to pick a balanced tree, deep enough to view all the features we chose before while not too deep, which can result in overfitting and high complexity.

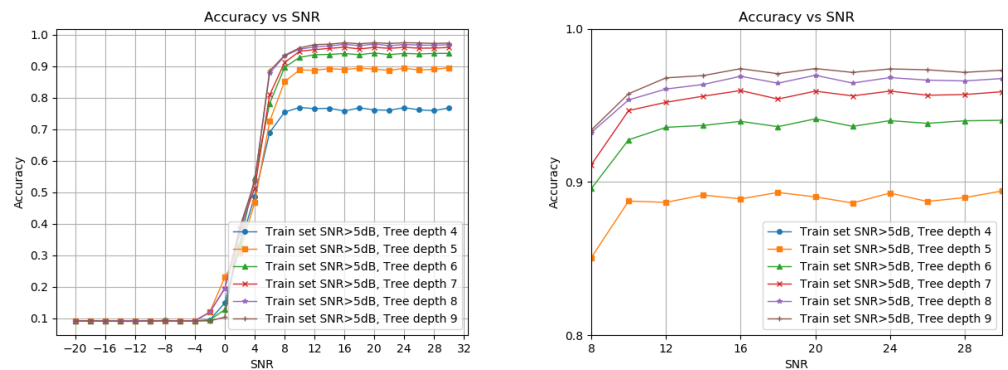


Figure 2. P_d based on tree depth. (a) Full view; (b) Zoomed in view.

Figure 2 shows the detection probability under different decision tree depths. We start at four layers, since this is the minimal depth for 11 labels, and stop at nine as we see no significant improvement from 8. The optimal depth for each application can be chosen based on memory and runtime requirements. We picked an 8-deep tree with a higher than 5 dB SNR train set.

We aim to compare our results to state-of-the-art deep learning models. First, we compare our method with three different state-of-the-art models of deep neural networks as examples. The models we examine are ResNet [22], DenseNet [23], LSTM [10], FLANs [8] and VGG [24]. We aim to compare our model to the neural networks in two manners. Figure 3 shows the first comparison of neural networks in the classification accuracy.

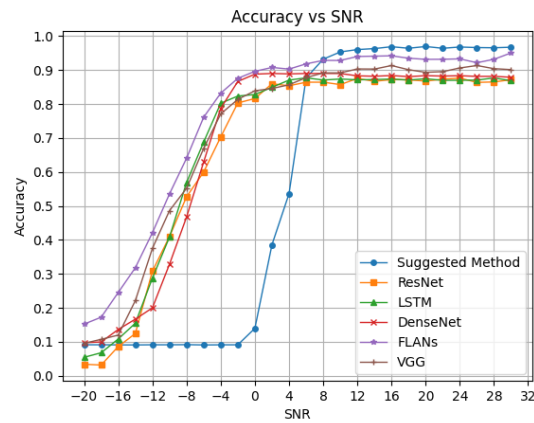


Figure 3. P_d comparison for the suggested method.

We can see the advantage of deep neural networks in the negative SNR zone, while our method fails in comparison. However, we have already discussed why classification at negative SNR is usually irrelevant in terms of perfect demodulation. Starting from a splitting threshold of 6 dB SNR, our suggested method yields better detection accuracy than the popular NN methods by a margin of almost 8%.

Table 2. Parameter size of different models.

Model	Total Parameters
6 Deep Tree	63
7 Deep Tree	127
8 Deep Tree	255
ResNet	214,922
LSTM	770,378
DenseNet	542,850
VGG	1,621,780
FLANs	954,747

The secondary comparison with the Neural Network models primarily centers on the parameters or memory requirements. We use Equation (15) to compute the total parameters necessary for diverse tree depths, subsequently contrasting these values in Table 2 alongside the proposed NN models.

Our model, even with the inclusion of additional layers, consistently maintains a parameter count ranging in the few hundred range. In contrast, the NN models register parameters in the hundreds of thousands. This substantial discrepancy highlights the sparse memory utilization inherent in our model, a characteristic that holds substantial significance. It's particularly advantageous when considering the amalgamation of multiple models to bolster accuracy, especially in scenarios necessitating low complexity and abbreviated runtime. This streamlined memory usage serves as a pivotal advantage, especially in contexts where computational efficiency and rapid processing are imperative.

Moreover, this substantial difference in parameter count underscores a critical limitation in NN models, notably their impracticality for deployment within SDR systems. The hefty memory requirements demanded by NN architectures, often in the realm of millions of parameters, render them impractical for many SDR implementations. In contrast, our model's ability to operate within the realm of mere hundreds or thousands of parameters renders it significantly more feasible for integration within SDR systems. This advantage positions our approach as particularly suitable for real-world deployment scenarios where

memory limitations are stringent, allowing for more practical utilization within SDR-based Cognitive Radio networks.

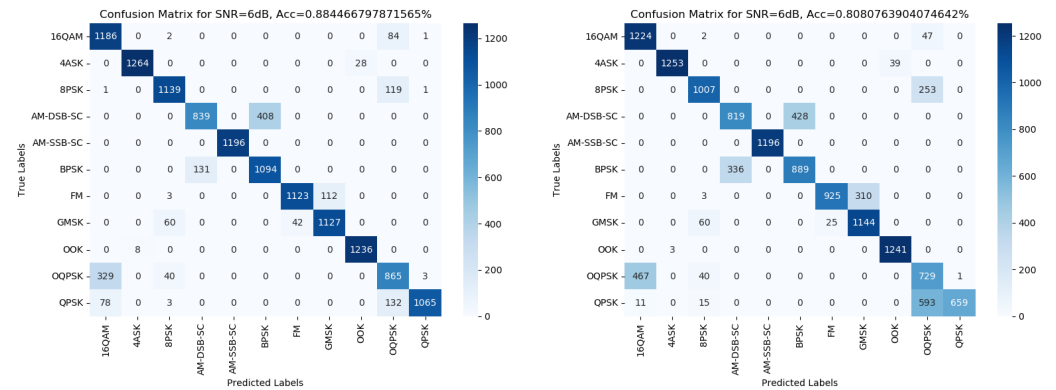


Figure 4. Confusion Matrices for different tree depth: (a) depth=7; (b) depth=8.

We examine the confusion matrix from the suggested method. Figure 4 shows the confusion matrices at 6 dB for the 7 and 8 deep decision trees under the 5 dB SNR threshold train set. We observe in the illustrated figures, as well as in other matrices we examined, that the mislabeling of the model is consistent – meaning the labels that can get mismatched are almost always the same. Based on this observation, we can split the modulation schemes into subgroups as follows:

- Group 0 - AM-SSB;
- Group 1 - AM-DSB and BPSK;
- Group 2 - 4ASK and OOK;
- Group 3 - FM and GMSK;
- Group 4 - 8PSK, 16QAM, QPSK and OQPSK.

Group 0 exclusively comprises SSB modulation due to its nearly instantaneous classification across an array of features, even at relatively low SNR) values. This consistent and rapid detection is observed consistently across various models tested, establishing SSB as a distinct and easily recognizable modulation type.

On the other hand, Groups 1, 2, and 3 collectively represent pairs of modulations that exhibit a tendency to be confused or mixed up, particularly prevalent at lower SNR levels. This occurrence suggests a noteworthy statistical similarity in the signal models of these modulation schemes. This statistical resemblance often results in misclassification or overlapping characteristics, leading to the assumption that the underlying signal models within these groups share comparable statistical behaviors, thereby complicating their differentiation, especially under lower SNR conditions.

Group 4, encompassing the remaining modulations, portrays a more intricate scenario where these modulations commonly intermingle without any particular pair exhibiting a significantly higher occurrence than the others. This lack of distinct differentiation signifies a higher level of complexity within this group, where multiple modulation schemes exhibit overlapping statistical features or behaviors, making their individual identification considerably more challenging, especially in contexts where their distinguishing characteristics are blurred due to similarities.

This categorization provides valuable insights into the behavior and detectability of various modulation groups, shedding light on the distinct characteristics observed across different SNR levels and establishing patterns of similarity or complexity among them.

4. Discussion

We conducted simulations to validate our hypothesis on the benefits of using a classic Feature-Based algorithm for AMC. Our initial focus was to examine the impact of individual features on detection accuracy, which led us to remove unnecessary or redundant features

during the model's training phase. We then partitioned the data into high and low SNR groups to demonstrate how specific modulation schemes become detectable only above a certain SNR threshold. This finding allowed us to refine our model by focusing on the essential SNR levels relevant to real-world systems. The final phase of our simulation aimed to fine-tune the tree's hyperparameters to ensure the highest accuracy while maintaining optimal complexity.

Next, we compared our finalized model against three NN models, which yielded varied results. On one hand, our proposed method exhibited superiority over NN models in terms of detection accuracy, especially beyond the specified threshold SNR values. On the other hand, our method struggled to detect signals with low SNR, where NN models excelled. However, as previously noted, many modulation schemes become irrelevant for demodulation below certain SNR thresholds. Apart from detection accuracy, we also compared our method with NN models in terms of memory storage. While basic NN models required millions of deep parameters, our straightforward classifier required a maximum of a few hundred parameters, depending on the chosen model. This advantage significantly impacted end units as the variance in total parameters directly influenced memory usage and runtime during the classification process. These factors directly contribute to improved battery life, reduced costs, and enhanced speed in desired end units.

However, our suggested method does have limitations. Primarily, the algorithm may not be optimal for discussions involving low SNR scenarios and modulations that are decodable even with an SNR lower than 2–4 dB. Additionally, we observed instances where certain modulation schemes were erroneously identified even at higher SNR. This occurrence might be attributed to similar carriers and statistics among some modulations, indicating that our method might not be suitable for those specific schemes. Despite these limitations, all outcomes affirm that the classic FB classifier for AMC can be both successful and advantageous when considering its application in real-life Cognitive Radio networks.

5. Conclusions

Our research focused on developing a simple yet effective method for automatic modulation classification in spectrum sensing. We found that our proposed method outperforms some of the most widely used neural networks in terms of both accuracy and memory efficiency. To achieve this, we chose the decision tree classifier for its simplicity, which allowed us to identify efficient statistical features like High-Order Cumulants (HOC) that are specifically tailored for each modulation scheme. Our study went beyond the usual emphasis on digital modulations and explored the possibility of using cumulants to classify analog modulations, which has not been extensively researched before. By identifying specific cumulants for each modulation type, we discovered that there are subsets of modulations that share similar statistical behaviors. This discovery not only helped us understand modulation characteristics better but also provided a foundation for improving classification methodologies. Our approach is optimal for end units within Cognitive Radio (CR) networks, as it achieves high accuracy with minimal parameters, which is crucial for memory and speed efficiency. In our future work, we plan to leverage these identified modulation subgroups to develop a more intricate method that further improves accuracy. We also aim to expand our approach to include a wider range of modulations, making our classification framework more versatile and applicable.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: This research is based on the published dataset by Timothy James O'Shea; Tamoghna Roy and T. Charles Clancy. Available at <https://www.deepsig.ai/datasets>.

Acknowledgments: Not applicable

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMC	Automatic modulation Classification
CR	Cognitive radio
SDR	Software defined radio
SNR	Signal to Noise Ratio
FB	Feature based
HOS	High order statistical moments
HOC	High order cumulants
AWGN	Additive white Gaussian noise

References

- Mitola, J. (1995). The software radio architecture. *IEEE Communications magazine*, 33(5), 26–38.
- Mitola, J., and Maguire, G. Q. (1999). Cognitive radio: making software radios more personal. *IEEE personal communications*, 6(4), 13–18.
- Haque, M., Sugiura, Y., and Shimamura, T. (2019). Spectrum Sensing Based on Higher Order Statistics for OFDM Systems over Multipath Fading Channels in Cognitive Radio. *Journal of Signal Processing*, 23(6), 257–266
- Bozovic, R., and Simic, M. (2019). Spectrum Sensing Based on Higher Order Cumulants and Kurtosis Statistics Tests in Cognitive Radio. *Radioengineering*, 29(2).
- Salahdine, F. (2017). Spectrum sensing techniques for cognitive radio networks. *arXiv preprint arXiv:1710.02668*.
- Callaghan, T. G., Perry, J. L., and Tjho, J. K. (1985). Sampling and algorithms aid modulation recognition. *Microwaves*, 24, 117–119.
- Nebabin, V. G., and Sergeev, V. V. (1984). *Methods and techniques of radar recognition*. Moscow Izdatel Radio Sviaz.
- Zhang, D., Lu, Y., Li, Y., Ding, W., Zhang, B., and Xiao, J. (2023). Frequency learning attention networks based on deep learning for automatic modulation classification in wireless communication. *Pattern Recognition*, 137, 109345
- Ma, J., Hu, M., Wang, T., Yang, Z., Wan, L., and Qiu, T. (2023). Automatic Modulation Classification in Impulsive Noise: Hyperbolic-Tangent Cyclic Spectrum and Multibranch Attention Shuffle Network. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–13.
- Elsagheer, M. M., and Ramzy, S. M. (2023). A hybrid model for automatic modulation classification based on residual neural networks and long short term memory. *Alexandria Engineering Journal*, 67, 117–128.
- Maroto, J., Bovet, G., and Frossard, P. (2023, June). Maximum Likelihood Distillation for Robust Modulation Classification. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1–5). IEEE.
- Dobre, O. A., Abdi, A., Bar-Ness, Y., and Su, W. (2007). Survey of automatic modulation classification techniques: classical approaches and new trends. *IET communications*, 1(2), 137–156.
- Hazza, A., Shoaib, M., Alshebeili, S. A., and Fahad, A. (2013, February). An overview of feature-based methods for digital modulation classification. In *2013 1st international conference on communications, signal processing, and their applications (ICCSPA)* (pp. 1–6). IEEE.
- Sun, X., Su, S., Zuo, Z., Guo, X., and Tan, X. (2020). Modulation classification using compressed sensing and decision tree-support vector machine in cognitive radio systems. *Sensors*, 20(5), 1438.
- Shih, P., and Chang, D. C. (2011, August). An automatic modulation classification technique using high-order statistics for multipath fading channels. In *2011 11th international conference on ITS telecommunications* (pp. 691–695). IEEE.
- Hazza, A., Shoaib, M., Alshebeili, S. A., and Fahad, A. (2013, February). An overview of feature-based methods for digital modulation classification. In *2013 1st international conference on communications, signal processing, and their applications (ICCSPA)* (pp. 1–6). IEEE
- Swami, A., and Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. *IEEE Transactions on communications*, 48(3), 416–429.
- Jdid, B., Lim, W. H., Dayoub, I., Hassan, K., and Juhari, M. R. B. M. (2021). Robust automatic modulation recognition through joint contribution of hand-crafted and contextual features. *IEEE Access*, 9, 104530–104546.
- Fryntov, A. E. (1989). Characterization of a Gaussian distribution by gaps in its sequence of cumulants. *Theory of Probability and Its Applications*, 33(4), 638–644.
- Volkmer, H. (2014). A characterization of the normal distribution. *J. Stat. Theory Appl.*, 13(1), 83–85.
- O’Shea, T. J., Roy, T., and Clancy, T. C. (2018). Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1), 168–179.
- Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7), 3142–3155.

23. Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700–4708).
24. Ghanem, H. S., Al-Makhlaway, R. M., El-Shafai, W., ElSabrouty, M., Hamed, H. F., Salama, G. M., El-Samie, F. E. A. (2023). Wireless modulation classification based on Radon transform and convolutional neural networks. *Journal of Ambient Intelligence and Humanized Computing*, 14(5), 6263-6272.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.