IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 July 2022, accepted 29 July 2022, date of publication 8 August 2022, date of current version 15 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3197224

SURVEY

Wireless Signal Representation Techniques for Automatic Modulation Classification

XUEYUAN LIU, CAROL JINGYI LI, (Student Member, IEEE), CRAIG T. JIN, AND PHILIP H. W. LEONG, (Senior Member, IEEE)
School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia
Corresponding authors: Xueyuan Liu (xueyuan.liu@sydney.edu.au) and Philip H. W. Leong (philip.leong@sydney.edu.au)

ABSTRACT In this paper, we present a comprehensive survey and detailed comparison of techniques that have been applied to the problem of identifying the type of modulation contained within received wireless signals. Known as automatic modulation classification (AMC), the problem has been studied for many decades. AMC plays a significant role in both military and civilian scenarios and is the main step in smart receivers. With the development of software-defined radios and automatic communication systems, IoT technology and the spread of 5G technology, there has been exponential growth in the number of spectrum-using equipment making the issue of scarce spectrum resources more prominent. Although AMC techniques can be optimized from the classifier's point of view, signal pre-processing also plays a critical role. Relevant data representation approaches include time-frequency analysis, cyclostationary transforms, and hybrid techniques. We provide a taxonomy of common approaches based on order and dimensionality along with an overall analysis of signal pre-processing algorithms for AMC. Furthermore, we reproduce the major existing schemes under uniform conditions, allowing an objective comparison among different methodologies. Finally, we create an open-source reproducible Python library to simulate these techniques, ensuring the usefulness for future research.

INDEX TERMS Modulation classification, signal pre-processing, high order statistics, cyclostationary, time-frequency transformation.

I. INTRODUCTION

With the recent advances in telecommunication technologies, user applications have become increasingly bandwidth-critical. Heterogeneous networks have evolved to include diverse types of data traffic based on end-user requirements to support existing voice and data services, and meet the next-generation network requirements. The control of spectrum resources requires real-time information extraction and processing within a context of expanding heterogeneity and dynamic systems. Automatic modulation classification (AMC) refers to the problem of identifying the modulation method of received radio signals from a specified set of possible modulation techniques. In implementation, it has been described as an intermediate procedure between signal detection and demodulation [1]. It plays a crucial role in various civilian, commercial and military scenarios.

AMC techniques have been classified into two broad methods: likelihood-based (LB) and statistical feature-based (FB) approaches [1]. The LB approaches show good performance by building a probabilistic model for the received signal, but suffer from high computational complexity and require knowledge of the signal environment and channel parameters in advance [2]. Some LB algorithms such as the generalized likelihood ratio test and the discrete likelihood function test have been proposed to reduce the complexity. Although LB algorithms can increase the performance efficiency by applying maximum likelihood estimations or multiple lookup tables [3], [4], they do not perform well on some modulation modes or other tasks [2].

FB techniques aim to extract salient features for AMC from the received time series [5]. If waveform features are capable of identifying the type of modulation and if classifiers can

The associate editor coordinating the review of this manuscript and approving it for publication was Vicente Alarcon-Aquino.

create suitable decision boundaries, then FB algorithms have the potential to achieve excellent performance with fewer resources [1], [6]. Most of the existing well-developed AMC techniques are FB-based, especially with the emergence of machine learning techniques such as deep neural networks (DNNs) [7], [8].

In this paper, we conduct an in-depth review of feature-based digital modulation recognition systems, e.g. [9]–[21], that embody the processing framework shown in Fig. 1. We use the term *pre-processing* to describe the feature extraction process which includes methods such as the fast Fourier transform (FFT), spectral correlation function (SCF), higher order statistics (HOS), etc. [22]–[24] and the term *classifier* to refer to the processing module which maps the features to an AMC class. Classifier modules include methods such as decision trees (DT), support vector machines (SVM), convolutional neural networks (CNN), recurrent neural networks (RNN), long short-term memory (LSTM), etc. These are reviewed in Section II-C. A taxonomy of pre-processing techniques for AMC is presented in Table 1. The pre-processing techniques are grouped into twelve different classes based on the order of the transforms involved and the dimensionality of the features. The original raw sequences in the first row of the table can be considered as an identity transformation. The focus of this study are the starred rows, i.e. the last three rows.

In this paper, the highlights are listed as follows:

- A comprehensive survey of pre-processing techniques and their applications in AMC that follows the taxonomy that groups previously published and is based on order and dimensionality.
- Objective tests and comparative analysis of existing dominant pre-processing techniques are conducted using the publicly available DeepSig RadioML dataset [25] and a single-layer Linear Discriminant Analysis (LDA) classifier as a benchmark. While many reviews of AMC techniques have been published [2], [6], [7], to the best of our knowledge this is the first quantitative and reproducible comparison.
- This article describes in detail the principles of various techniques, and compares their advantages and disadvantages, including computational complexity and application scenarios, which provides a valuable reference for future blind recognition studies.
- An open-source Python 3 [26] library which implements all the techniques presented in this study is explained in Section VII-A.

The remainder of this paper is organized as follows. In Section II we review digital modulation and classifier techniques. In Section III, linear time-frequency techniques are reviewed. Section IV explains the higher order statistics methods. Section V summarizes all of the cyclostationary-based algorithms including second and higher order spectral correlation functions and kernel-based techniques. Section VI reviews and discusses the prominent AMC works using data pre-processing. Section VII describes the replication experiments and objective comparative analysis of the afore-mentioned techniques. Finally, conclusions are drawn in Section VIII.

II. BACKGROUND

This section reviews basic models for wireless signals, the universal transformation framework, cyclostationary analysis as a pre-processing scheme, popular AMC classifiers, and other topics.

A. DIGITAL SIGNAL MODULATION

Digital modulation refers to the process of encoding digital information using the amplitude, phase or frequency of a transmitted signal. The choice of encoding method influences the achieved data rate and robustness to noise in the channel. Typically, modulation techniques encode several bits into a single symbol, and the rate at which each symbol is transmitted determines the bandwidth of the transmitted signal [9].

The baseband complex envelope of a continuous-time digitally modulated signal is:

$$x_L(t) = \sum_n s[n]g(t - nT) \tag{1}$$

where $s[n] \in \mathbb{C}$ are numbers in the complex plane, T is the symbol period, and $g(\cdot)$ is a pulse shaping function, e.g. a raised-cosine shaping filter [10]. The signal $x_L(t)$ can be orthogonally decomposed into $I(t) + jQ(t)$, where $I(t) = Re[x_L(t)]$ is referred to as the in-phase component and $Q(t) = Im[x_L(t)]$ as the quadrature component. The baseband real signal is typically modulated with a carrier to obtain the transmission signal: $x(t) = Re\{x_L(t)e^{j\omega_c t}\} = I(t)\cos(\omega_c t) - Q(t)\sin(\omega_c t)$, where $\omega_c = 2\pi f_c$ and f_c is the carrier frequency. Its equivalent complex analytic signal is: $z(t) = x_L(t)e^{j\omega_c t} = x(t) + j\hat{x}(t)$, where $\hat{x}(t)$ is the Hilbert transform of $x(t)$.

The aim of the receiver is to recover $x_L(t)$, however, hardware differences between the receiver and transmitter as well as channel effects make this challenging. It is crucial to match the receiver's carrier frequency with that of the transmitter. This is not perfectly possible in general, resulting in a carrier frequency offset. In addition, fading within the channel can also occur. In this light, the decoded baseband signal at the receiver is described by the following signal model (note that we assume an additive Gaussian white noise (AWGN) channel):

$$\tilde{x}_L(t) = a \cdot e^{-j(2\pi \Delta f t + \varphi)} \sum_n s[n]g(t - nT - N_1) + w(n) \tag{2}$$

where a is channel gain, Δf is the carrier frequency offset, φ is the channel phase (which also includes the carrier phase offset), N_1 is a timing offset, and $w(n)$ is the channel noise.

In this review, we focus on digital modulation techniques, such as OOK (On-off keying), 4ASK (Amplitude shift keying), 8ASK, BPSK (Binary Phase shift keying), QPSK

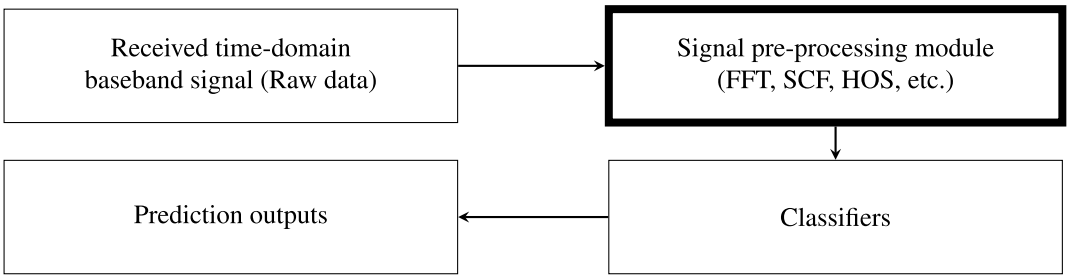


FIGURE 1. General flow chart for pre-processing based AMC.

TABLE 1. A taxonomy of AMC feature extraction techniques.

	Scalar features	1D features	2D features
Original raw sequences	• Standard deviation of the absolute value of amplitude or phase	• IQ series	• Constellation graphs
*** Time-frequency transformations (FT, WT, ST)**	• Peak values and positions • Number of transforms' peaks • Specific correlation coefficients and standard deviations of transforms	• Fourier transform family: FFT, DFT, STFT • Wavelet transform family: DWT, CWT	• STFT image, time axis vs frequency axis • S-transform spectrum • Wigner distribution function
*** Second order cyclostationary (SCD, CCSD, HTC)**	• Specific axis's peak counts or positions (among α or f profile)	• α profile • f profile • Specific frequency profile	• CAF/CCF images, α vs τ • SCD/CCSD/HTC images, α vs f
*** High order cyclostationary (HOS, CTCF)**	• C_{mn} • M_{mn} • RD-CTCF$_{mn}$ value at specific α	• RD-CTCF$_{mn}$ sequences	• High order cyclic polyspectrum

(Quadrature-PSK), 8PSK, 16PSK, 32PSK, 16APSK (Amplitude Phase shift keying), 32APSK, 64APSK, 128APSK, 16QAM (Quadrature Amplitude modulation), 32QAM, 64QAM, 128QAM, 256QAM, GMSK (Gaussian Minimum-shift keying), and OQPSK (Offset Quadrature Phase-shift keying).

B. PRE-PROCESSING AND CYCLOSTATIONARY FEATURES

Pre-processing approaches can be characterized by their statistical order, i.e., the power of a sample used in statistical functions. Common pre-processing methods include linear transforms like the fast Fourier transform and wavelet transform. These only involve low-order statistical characteristics, limiting their ability to distinguish between modulation types. In contrast, the cyclostationary features have many advantages as a signal pre-processing method in FB classifiers, and can be used to explore the underlying properties of the signal. This is because most actual signals in real scenarios show uncertainty in direct observation.

A cyclostationary process is a signal having statistical properties that vary cyclically with time. It was first analyzed to investigate the diversity in the spectrum appearance of all modulation types in 1994 [27]. Various work have confirmed that the cyclic properties of a signal are exhibited at second and higher statistical orders, e.g., using measures such as spectral correlation density, high order moments and cumulants, cyclic statistics, as well as kernel-based cyclic correlation functions. Significantly, each modulation type has a different periodic spectral pattern [28], [29]. HOS analyses are often more robust to Gaussian and impulsive noise, and channel interruption. As well, HOS-based methods remain relatively straightforward to calculate, for example, the Taylor expansion of the Gaussian kernel in the correntropy method has high order terms embedded within the expansion.

Due to the flexibility and developability of cyclostationary functions, there are many frameworks based on these functions, but their general format for cyclic frequency function can be written as:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{n=-T/2}^{T/2} Y^{(*)}[n,\underline{\tau}]e^{-j2\pi\alpha n} \tag{3}$$

Here, $\underline{\tau}$ denotes individual delays $(\tau_1,\ldots,\tau_m)$, m is customizable, for example, when the cyclic temporal cumulants function (CTCF) is involved, it relates to different terms with varying time delays. The notation $^{(*)}$ indicates optional complex conjugation, and α is the cyclic frequency. Note that the definition of $Y[n,\underline{\tau}]$ is not standardized, so we let it be a customized function of the received signal $x[n]$. Many studies target cyclostationary features by developing the $Y^{(*)}[n,\underline{\tau}]$ in (3), as summarized in Table 2. Different algorithms exhibit varying advantages with respect to additive white Gaussian noise (AWGN), impulsive noise, and channel interference [10], [30], [31]. The channel interference does less corruption to the feature maps obtained from these operations.

TABLE 2. Design of $Y[n, \underline{\tau}]$ in (3) (Note: specific formula descriptions are in later sections).

Description Cyclic Y Function	Expression $Y[n, \underline{\tau}]$
Autocorrelation	$x[n]x^*[n+\tau]$
Correntropy	$G_\sigma(x[n]x^*[n+\tau])$
Hyperbolic-tangent	$\tanh(x[n]x^*[n+\tau])$
High-order cumulants	$C_{p,q}(x^{p-q}[n+\underline{\tau}](x^*)^q[n+\underline{\tau}])$

C. CLASSIFIERS FOR MODULATION RECOGNITION

Recent high-performance classifiers apply machine learning models [32]. In this work, we consider classifiers appropriate for AMC. These can be divided into supervised, unsupervised, and deep models. For an efficient implementation, the choice of model should relate to the specific input data.

1) LDA AND PCA

Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) are commonly used linear supervised and unsupervised techniques respectively [33]. Both approaches enable classification based on a low-dimensional embedding of the modulation data. PCA utilizes the data's covariance or correlation matrix to determine the directions of maximal variability - these correspond to the eigenvectors with the largest eigenvalues. Although LDA is also trained to compute the top largest eigenvalues' eigenvectors, it works for class separation through additional intra- and inter-class weight matrix operations.

2) DECISION TREE

A DT is a supervised multi-class model which uses a tree graph to make classification decisions at tree nodes based on feature values. Training primarily of DT is to design the decision criteria for each tree node and to set the correct threshold values. Various studies used HOS as the features for DT [34]–[36], while some combined the instantaneous signal and HOS features for training DT [37].

3) SUPPORT VECTOR MACHINE

Support vector machines uses a kernel function to project the data into a high dimensional space after which classification in then based on hyperplanes. For AMC, SVM has good performance at various SNR ranges [38]–[40] and can overcome the overfitting problem of deep networks [29]. However, these are generally limited to a small number of training categories, and the results will be restricted if high order modulations such as 256QAM are included. As well, SVM needs assistant schemes to do multi-classification such as kernel function and high-dimension feature space, these schemes incur increased computational cost.

4) NEURAL NETWORK

The typical representation of deep model, neural networks, are comprised of connected computational nodes referred to as neurons. Each neuron computes a weighted sum of its inputs followed by a nonlinear activation function. The neurons are organised into layers that progressively feed forward from input to output. Training determines the optimal weights based on the training data. Deep network methods for AMC commonly rely on well-defined modules referred to as convolutional neural networks (CNN) and recurrent neural networks (RNN) based on gated recurrent units (GRU), and long short term memory (LSTM) units [41]. Many studies use customised networks for AMC, while the most common and best performing standard networks are ResNet and AlexNet [42]. Training deep neural networks require considerable computational resources and labelled data sets. Thus techniques for reducing the training time and computational complexity of NNs are helpful. For example, principal component analysis (PCA) can be utilized to compress the data and improve the classification accuracy at low SNR [43].

III. LINEAR TIME-FREQUENCY ANALYSIS

Since most of the signals of interest are non-stationary, a number of techniques have been developed to describe signals with dynamic contents, such as short time Fourier transform, wavelet transform and S-transform. These are the basis of the field of time-frequency analysis.

A. FOURIER TRANSFORM

As the most classical fundamental transform, the Fourier transform of $x[n]$ has the following equations in the discrete domain:

$$\mathrm{DFT}(x) \equiv X(k) = \sum_{n=0}^{N_s-1} x[n] e^{-j\frac{2\pi}{N_s}nk} \tag{4}$$

$$\mathrm{STFT}(x)(\tau, k) \equiv X(\tau, k) = \sum_{n=0}^{N_s-1} x[n]h[n-\tau] e^{-j\frac{2\pi}{N_s}nk} \tag{5}$$

here, τ denotes a time delay, N_s is the number of sample points, $h(\cdot)$ is a window function. The application of FT or DFT in AMC is often combined with other features, in general, the most common features include calculating the number of peaks ζ and the maximum value γ of spectral power density of the normalised centred instantaneous amplitude, as shown in (6) and (7) [44], [45].

$$\zeta = \mathrm{Number}\{|DFT(x)| > \epsilon\} \tag{6}$$

$$\gamma = \frac{\max |DFT(a_{cn}(n))|^2}{N_s} \tag{7}$$

a_{cn} denotes the normalised centred instantaneous amplitude of the signal, N_s is the number of sample points. Since the modulation techniques M-PSK, M-QAM, and M-FSK have different numbers of frequency peaks, the peak count of the magnitude of the DFT can be used as a feature. For example, the ϵ in (6) is customized to 0.7 times the maximum value so that ζ can be used to distinguish M-FSK [45].

In (5), the non-stationary signal can be regarded as a superposition of a series of short-time stationary signals,

highlighting the variation of the original signal frequency with time delay. The STFT decomposes the time domain signal into windowed segments and then applies the discrete Fourier transform on each segment. For this consideration, the standard deviation between the peaks of absolute value after each window-STFT can be used for identification [45], this deviation is shown in (8).

$$\zeta = \sqrt{\frac{1}{N_w - 1}\sum_{i=1}^{N_w-1}[W_i - \overline{W}]^2}$$
$$\text{where } W_i = max|STFT_{i+1}| - max|STFT_i| \quad (i = 0, 1, \ldots, N_w - 1) \tag{8}$$

N_w represents the number of windows, W_i is the difference in peak values between adjacent windows. Additionally, the number and size of peaks can be calculated separately for the time-shift axis or the frequency axis as features. The spectrogram corresponding to the STFT can be used as an input image for a CNN [46].

In practice, each FT only transforms a finite window of time-domain data and therefore, signal truncation is required. For periodic signals, if the truncation is not an integer multiple of the period (period truncation), then there will be a leakage in intercepted signal [47]. In order to minimize this error, a weighting function (window function) can be applied. In addition to the type of window, the window length and hop size are important parameters that impact the window's performance. When the signal's frequency content or channel characteristics change rapidly with time, a smaller window size is needed. More generally, different window sizes should be tried to obtain optimal performance.

FT features can give better results if they are combined with other features [48]. In particular, the Fourier transform can be beneficial if the data set involves different frequency spectrum characteristics in terms of modulation schemes or bandwidth, for example, discriminating different types of M-FSK [49]. However, for broad AMC applications with various modulations, the performance of FT features may sometimes not be as good as directly using the original IQ sequences [50].

B. WAVELET TRANSFORM

The wavelet transform (WT) [51] addresses the FT's limitations in capturing features from burst signals, non-stationary processes and transient changes [52], [53] and hence the WT can effectively analyze waveforms with sharp discontinuities and peaks. Moreover, some fast algorithms derived from the WT (e.g., the Mallat algorithm [54]) are computationally inexpensive and can perform real-time identification.

The WT can be described as the projection of a signal based on scaled and time-shifted versions of the original wavelet (mother wavelet) [52]. The benefit of wavelet analysis lies in its scale-time view of a signal and its ability to enable multi-resolution analysis [55]. The standard continuous wavelet transform (CWT) of a received complex signal is given by:

$$CWT(a,\tau) = \frac{1}{\sqrt{|a|}}\int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{a}\right)dt \triangleq \frac{1}{\sqrt{|a|}}\sum_n x[n]\psi^*\left[\frac{n-\tau}{a}\right] \tag{9}$$

where a is the scale factor, τ is the translation variable, $\psi(\cdot)$ is the wavelet generating function, and the asterisk denotes complex conjugation. The discrete wavelet transform (DWT) directly follows from the CWT, sampling the scale and translation parameters of the continuous wavelet [56]:

$$DWT(j,k) = \sum_n x[n]\psi_{j,k}[n]$$
$$\text{where } \psi_{j,k}[n] = \frac{1}{\sqrt{2^j}}\psi\left[\frac{n-k2^j}{2^j}\right] \tag{10}$$

here the amplitude of the scale transform does not have to be in the form of the nth power of 2, but this is commonly used for the convenience of calculation.

In AMC, most research works employ the CWT with the Haar wavelet, as the magnitude of the Haar WT can identify M-PSK, M-QAM, MASK, and M-FSK modulated signals [57], [58]. The Haar wavelet is defined by [52]:

$$\psi(t) = \begin{cases} 1 & 0 \le t \le \frac{T}{2} \\ -1 & \frac{T}{2} \le t \le 1 \\ 0 & otherwise \end{cases} \tag{11}$$

The Haar WT reveals the difference in amplitude, phase and frequency of the aforementioned digital modulations:

$$|CWT|_{ASK}(a,\tau) = \frac{4\sqrt{S_i}\sin^2\left(\frac{\omega_c a T_s}{4}\right)}{\sqrt{a}\omega_c} \tag{12}$$

$$|CWT|_{FSK}(a,\tau) = \frac{4\sqrt{S_i}\sin^2\left(\frac{(\omega_c+\omega_i) a T_s}{4}\right)}{\sqrt{a}(\omega_c+\omega_i)} \tag{13}$$

$$|CWT|_{PSK}(a,\tau) = \frac{4\sqrt{S_i}\sin^2\left(\frac{\omega_c a T_s}{4}\right)}{\sqrt{a}\omega_c} \tag{14}$$

$$|CWT|_{QAM}(a,\tau) = \frac{4\sqrt{S_i}\sin^2\left(\frac{\omega_c a T_s}{4}\right)}{\sqrt{a}\omega_c} \tag{15}$$

where a is the scale factor, T_s is the duration of the pulse shaping function, $S_i = I_i^2 + Q_i^2$ associates with the modulation modes, which is a constant for FSK, ω_c is the carrier frequency, and ω_i relates to the FSK configuration.

$|CWT|$ is almost constant for PSK signals, with peaks occurring at phase shifts. However, for ASK, FSK and QAM, their Haar WT plots look like a multi-step function due to variations in frequency or amplitude. To solve this problem, amplitude normalization can be performed before applying the WT, which distinguishes the FSK signal as it has the same HWT diagrams before and after normalization. Nevertheless, this is still not sufficient to identify QAM and ASK. One solution lies in choosing the optimal scale

factor. Wei and Cao [59] used the number of peaks in the histogram of the WT to quickly recognize ASK and QAM, which has high accuracy at low SNR. On the other hand, the statistical variance of the amplitude normalized WT can distinguish between ASK, FSK, PSK and QAM [57], [58], [60]. A summary of the main works on applying WT to AMC is given in Table 3.

Due to its multi-resolution analysis capability, WT is effective in extracting features from transient signals. It is therefore also used in a wide range of applications in image and signal processing, medicine, mathematics and other fields [61]. For example, WT can be applied to image compression and denoising [62], or signal denoising through wavelet decomposition, processing wavelet coefficients and signal reconstruction, and it can also achieve fault detection and localization of military equipment [63].

C. S-TRANSFORM

Like other time-frequency analysis tools, the S-transform (ST) can exhibit the energy distribution in both the time and frequency domains [67]. The notable feature is that it maintains the global reference as Fourier transform but can be considered a deformation of the wavelet transform as it has different resolutions at different frequencies [68]. The general S-transform is defined by:

$$s(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-((\tau-t)^2 f^2/2)} e^{-j2\pi ft} dt \qquad (16)$$

The S-transform can be a generalized form of STFT with a Gaussian window. Note that the Gaussian window also depends on the parameter f, which makes the window width vary with frequency. Thus, the S-transform is a time-frequency analysis with high frequency resolution at low frequencies, and high time domain resolution at high frequencies.

$$IF(\tau, f) = \frac{1}{2\pi} \times \frac{\partial \Phi(\tau, f)}{\partial \tau} \qquad (17)$$

In the work [69], the instantaneous value of ST is expressed in terms of amplitude A and phase Φ, for which the instantaneous frequency is obtained according to (17). In addition, the correlation coefficients between IF and τ, the variance of normalized IF are used as the extracted signal features for classification. Although the final accuracy is high, about 97.25% at SNR = 3dB, there are limited modulation types involved, only CP, LFM, BPSK and BFSK. The study [70] focuses on comparing the performance of WT and ST and applying different classifiers such as NN, Naive Bayes, LDA, SVM, k-NN to classify based on these features. The obtained results show that ST outperforms WT in terms of accuracy for all these classifiers and consumes much less time than WT. However, the shortcomings are, firstly, that the modulation types are still insufficient, only BPSK, QPSK, FSK and MSK, and the classifier parameter settings are not indicated. The literature [71] considers the ability of the ST for signal recognition in multiple scenarios, for example, AWGN, multipath Rayleigh fading channels, and underwater communication. The basic idea is first to perform the transformation, calculate its energy entropy according to the definition of information entropy, and input the features into an SVM classifier. The considered modulations include 16QAM, QPSK, 4FSK, OFDM, FM and SSB; although the results show that ST has good resolving power, direct comparisons with other techniques were not offered.

IV. SIGNAL STATISTICS

Early traditional features included variance of the amplitude, phase and frequency of normalized signals [72]. The important AMC works based on received signals' moments, cumulants and cyclic cumulants (CCs) are briefly reviewed here. The most representative and influential algorithm for digital modulation identification based on higher-order statistics, proposed by Swami and Sadler [73], uses the fourth-order cumulants of ideally synchronized and power normalized signals as classification features for BPSK, QPSK, 4PAM and 16QAM signals. The effects of SNR and a number of samples on recognition performance are also discussed. Marchand *et al.* [74] proposed a combination of fourth- and second-order CC amplitudes for QPSK and QAM classification. Spooner [75] utilized features of sixth-order CCs for PSK and QAM. The use of up to eighth-order CCs for a wider range of signals was studied in [13]. HOS is also shown to be useful in parameter estimation for non-stationary signals [76].

A. ESTIMATION OF MOMENTS AND CUMULANTS

High order statistics (HOS) refers to higher order moments and cumulants. The kth order moments of a set of random variables $x = [x_1, x_2, x_3, \ldots, x_k]^T$ are defined as the coefficients of $\omega = (\omega_1, \omega_2, \ldots, \omega_k)$ in the Taylor series expansion (if it exists) of the moment-generating function [77]:

$$\begin{aligned} m_{x_1, x_2, \ldots, x_k} &= E\{x_1 x_2 \ldots x_k\} \\ &= (-j)^k \frac{\partial^k \Phi(\omega_1, \omega_2, \ldots, \omega_k)}{\partial \omega_1 \omega_2 \ldots \omega_k}|_{\omega_1 = \ldots = \omega_k = 0} \end{aligned} \qquad (18)$$

where

$$\Phi(\omega_1, \omega_2, \ldots, \omega_k) = E\{\exp j(\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_k x_k)\} \qquad (19)$$

is the joint characteristic function of vector x, and $E(\cdot)$ is the expected value operator. If $x[n]$ is a complex-value time-series then its kth order moment can be calculated by: $m_{kq}(\tau_1, \ldots, \tau_{k-1}) = E(x[n] \ldots x[n + \tau_{k-q-1}] x^*[n + \tau_{k-q}] \ldots x^*[n + \tau_{k-1}])$ Here, k is the total order, q is the number of conjugate parts, and * means conjugation operation. Specifically, when $\tau_1 = \tau_2 = \ldots = 0$, this becomes the kth order raw moment of a random cyclostationary process and the moment formula becomes:

$$m_{kq}(x) = E(x^{k-q}[n](x^*)^q[n]) \qquad (20)$$

TABLE 3. Summary of WT-based AMC works.

Citation	Modulation Set	Classifier Design	Evaluation
[59]	M-QAM(M=16,64,128), M-ASK(M=2,4,8)	Using Haar wavelet function $\|CWT\|$ and the best scale factor. Also utilizes the number of peaks in histogram to realize fast classification.	This work finds the most appropriate scale factor for classification, and has high accuracy at low SNR. But this is restricted to AM signals.
[60]	QAM, PSK and FSK	Calculates normalized and unnormalized Haar $\|CWT\|$ variance and kurtosis, and then considers distribution of values to determine thresholds.	This method has a recognition rate of 97% and above when the SNR is greater than 5dB, even if SNR is lower than 5dB, the average accuracy can be still higher than 65%. Computational complexity is low.
[64]	BASK, BPSK, BFSK	Haar $\|CWT\|$ of signal is compared with specific modules containing transient information in wavelet domain to measure the similarity and calculate the cross- correlation between them.	The overall accuracy can reach 97.34% when SNR is -10dB, and when SNR is greater than -1dB the accuracy is almost 100%. Range of classes is limited (only for binary modulated signals).
[65]	2ASK, 4ASK, 8ASK, 16QAM, 32QAM, 64QAM, BFSK, QFSK, 8FSK, BPSK, QPSK and 8PSK	Classifier uses a tree-like structure. Each decision node uses the $\|CWT\|$ value combined with other statistical features. The four major modulation types are first classified, and then the modulation order M is decided.	The overall accuracy can reach 90% when SNR is zero, finding larger scale factors work better for this task. But only considers AWGN as channel interference.
[66]	2ASK, 4ASK, 8ASK, 2FSK, 4FSK, 8FSK, 2PSK, 4PSK, 8PSK and 16QAM	Constructs a coefficient of variation formula and a similarity measure function based on the difference of wavelet amplitude envelopes before and after normalization. MASK, M-QAM and M-FSK, M-PSK can be distinguished after normalization. Then, the changes of scale factor and coefficient of variation are used for secondary identification, and finally intra-class classification achieved using high order moments.	This study is relatively new for WT-based methods, so it combines the advantages of previous works. This algorithm is highly insensitive to the choice of parameters and has a wide range of applications. Also suitable when SNR is higher than 2dB. However, there are still limitations on the modulation orders.

The raw moments of signal are commonly used as an estimation of signal's moments, e.g. M_{40} stands for $m_{40}(x) = E(x^4[n])$. By default, $\tau_1 = \tau_2 = \ldots = 0$ and the subscripts represent the total order and the conjugation order.

Similarly, the kth order cumulant of process $x = [x_1, x_2, x_3, \ldots, x_k]^T$ can be derived from (18) using the cumulant generating function $\Psi(\omega_1, \omega_2, \ldots, \omega_k) = ln\Phi(\omega_1, \omega_2, \ldots, \omega_k)$:

$$\begin{aligned} c_{x_1,x_2,\ldots,x_k} &= cum(x_1, x_2, \ldots, x_k) \\ &= (-j)^k \frac{\partial^k ln\Phi(\omega_1, \omega_2, \ldots, \omega_k)}{\partial\omega_1\omega_2\ldots\omega_k}|_{\omega_1=\ldots=\omega_k=0} \end{aligned} \tag{21}$$

where there exists a transformation equation between high order moments and high order cumulants:

$$\begin{aligned} m_x(I) &= \sum_{\bigcup_{p=1}^{q} I_p=I} \prod_{p=1}^{q} c_x(I_p) \\ c_x(I) &= \sum_{\bigcup_{p=1}^{q} I_p=I} (-1)^{(q-1)}(q-1)! \prod_{p=1}^{q} m_x(I_p) \end{aligned} \tag{22}$$

A partition of a set X is a set of non-empty subsets of X such that every element x in X is in exactly one of these subsets [78]. Here $I = (1, 2, \ldots, k)$ is the indices of set $[x_1, x_2, \ldots, x_k]$, I_p is the unordered collection of set I, q is the number of partitions of I_p, the summation operator denotes the inner summation over all partitions of I.

Based on the above properties, cumulants can be estimated through moments (20), common relational expressions are given as follows:

$$\begin{aligned} C_{20} &= cum(x_n x_n) = M_{20} \\ C_{21} &= cum(x_n x_n^*) = M_{21} \\ C_{40} &= cum(x_n x_n x_n x_n) = M_{40} - 3M_{20}^2 \\ C_{41} &= cum(x_n x_n x_n x_n^*) = M_{40} - 3M_{20}M_{21} \\ C_{42} &= cum(x_n x_n x_n^* x_n^*) = M_{42} - |M_{20}|^2 - 2M_{21}^2 \\ C_{60} &= cum(x_n x_n x_n x_n x_n x_n) = M_{60} - 15M_{20}M_{40} - 30M_{20}^3 \\ C_{61} &= cum(x_n x_n x_n x_n x_n x_n^*) = M_{61} - 5M_{21}M_{40} - 10M_{20}M_{41} \\ &\quad - 30M_{20}^2M_{21} \\ C_{62} &= cum(x_n x_n x_n x_n x_n^* x_n^*) = M_{62} - 6M_{20}M_{42} - 8M_{21}M_{41} \\ &\quad - M_{22}M_{40} + 6M_{20}^2M_{22} + 24M_{21}^2M_{20} \\ C_{63} &= cum(x_n x_n x_n x_n^* x_n^* x_n^*) = M_{63} - 9M_{21}M_{42} + 12M_{21}^3 \\ &\quad - 3M_{20}M_{43} - 3M_{22}M_{41} + 18M_{20}M_{21}M_{22} \end{aligned} \tag{23}$$

Ideally, if a period of normalized digital modulated signal is sufficient long and contains all symbol points, then its theoretical HOS are displayed in Table 4.

B. HOS AS PRE-PROCESSING FEATURES

HOS or their variants can be directly used in AMC. Due to the symmetry of some digital modulated signal constellations, the nth order moments and cumulants for odd order are zero,

TABLE 4. Some typical digital modulation's HOS [27], [79].

	C_{40}	C_{41}	C_{42}	C_{60}	C_{61}	C_{62}	C_{63}	C_{80}	C_{81}	C_{82}	C_{84}
4ASK	-1.36	-1.36	-1.36	8.32	8.32	8.32	8.32	-111.85	-111.85	-111.85	-111.85
8ASK	-1.24	-1.24	-1.24	7.19	7.19	7.19	7.19	-92.02	-92.018	-92.02	-92.02
BPSK	-2	-2	-2	16	16	16	16	-272	-272	-272	-272
QPSK	1	0	-1	0	-4	0	4	-34	0	34	-34
8PSK	0	0	-1	0	-4	0	4	-34	0	34	-34
16PSK	0	0	-1	0	0	0	4	1	0	0	-34
16QAM	-0.68	0	-0.68	0	2.08	0	2.08	-14	0	-14	-14
32QAM	-0.19	0	-0.69	0	0.57	0	2.11	-2	0	-3.84	-13.8
64QAM	-0.62	0	-0.62	0	1.8	0	1.8	-11.5	0	-11.5	-11.5
128QAM	-0.18	0	-0.66	0	0.53	0	1.96	-1.8	0	-1.8	-12.5
256QAM	-0.6	0	-0.6	0	1.73	0	1.73	-11	0	-11	-11

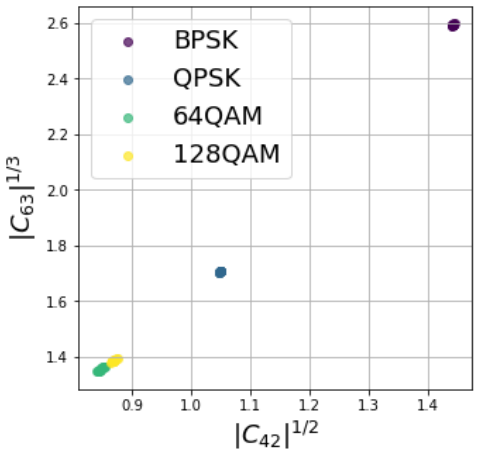


FIGURE 2. Normalized $|C_{42}|$ and $|C_{63}|$ of BPSK, QPSK, 64QAM and 128QAM under SNR +20dB.

so only their even order statistics are of interest [27], [80].

$$\begin{aligned} c_{p,q}(x_n) &= cum\Big(\underbrace{s_n e^{j2\pi f_o n}, \ldots, s_n e^{j2\pi f_o n}}_{p-q} \\ &\qquad \times \underbrace{s_n^* e^{j2\pi f_o n}, \ldots, s_n^* e^{j2\pi f_o n}}_{q}\Big) \\ &= \frac{1}{N}\sum_n e^{j2\pi(p-2q)f_o n} s_n^{p-q}(s_n^*)^q \\ &\triangleq \frac{C_{p,q}^{cst}}{N}\sum_n e^{j2\pi(p-2q)f_o n}\sum_m p\,(n-mT) \end{aligned} \quad (24)$$

For example, Fig. 2 shows $|C_{42}|$ and $|C_{63}|$ as features to discriminate between BPSK, QPSK, 64QAM and 128QAM. Here the normalization method introduced in [12] is adopted. Initially the differences in distribution between BPSK and QPSK are clear, while for 64QAM and 128QAM only using $|C_{42}|$ and $|C_{63}|$ is not sufficient. Nevertheless, HOS shows good ability to separate non-Gaussian signals from Gaussian noise for the analysis of chaotic systems [81].

For scalar HOS, issues arise when not all the symbols are represented in the processed signal [17], especially for high order modulation modes such as 256QAM, there is a high potential that only a small fraction of the constellation points are present in the received signal. In addition the carrier frequency offset, caused by the mismatch of receiver and transmitter, can also harm the scalar HOS. Several studies have investigated the impacts of this offset [17], [19], [80]. For the baseband digital modulated signals, their formats are given in (24), where $C_{p,q}^{cst}$ denotes the cumulant associated with symbol points s_n,f_o is the frequency offset, and N represents the total number of sample points. Though the term $e^{j2\pi(p-2q)f_o n}$ doesn't involve the magnitude, it affects the phase. This phase rotation will result in the term $\sum_n e^{j2\pi(p-2q)f_o n} \approx 0$, thereby causing the final value of time-averaged cumulant $c_{p,q}(x_n)$ close to zero when $p \neq 2q$ or carrier frequency offset f_c is not zero. Ideally, the numerical values of $C_{p,q}^{cst}$ and $c_{p,q}(x_n)$ should be equal, and $c_{p,q}(x_n)$ has been abbreviated as C_{pq} in most studies. Various techniques have been proposed to circumvent these problems, improving the classification capabilities of HOS:

- Multiply an additional offset compensation term $e^{-j2\pi(p-2q)f_o n}$. This only works when the carrier frequency is known or can be estimated [19].
- For each time n only use the absolute value of cumulants, for example: $\sum_n |e^{j2\pi(p-2q)f_o n} s_n^{p-q}(s_n^*)^q|$. Note that this may lead to different results from theoretical values as calculating absolute eliminate the negative numbers.
- Directly apply the absolute value $|C_{pq}|$ [17], [82], [83].
- Normalization methods [14], [15], [20] achieved by dividing different orders' cumulants, for instance:

$$\frac{|C_{42}^m|}{|C_{21}^n|}, \frac{|C_{63}^m|}{|C_{21}^n|}, \frac{|C_{63}^m|}{|C_{42}^n|} \quad (25)$$

  Here, m, n are customized power of cumulants according to the specific situation, which can be used in conjunction with the cumulant's order to dissolve the impacts of channels and signal power as well.

With regard to current practice, it is most common to select the representative normalized elements of HOS and use them as inputs to a DNN-based model (e.g. fourth-, sixth- and eighth-order cumulants such as $|C_{40}|$, $|C_{41}|$, $|C_{42}|$, $|C_{60}|$, $|C_{61}|$, $|C_{62}|$, $|C_{63}|$, $|C_{80}|$, $|C_{84}|$), which are summarised in Section VI and Section VII-B.

V. CYCLOSTATIONARY ANALYSIS

A. SECOND ORDER: CAF AND SCD

While the FT and WT can be classified as first-order statistical characteristics, autocorrelation $R_{xx}[\tau]$ and cross-correlation $R_{xy}[\tau]$ are second-order. A signal is wide sense almost-cyclostationary of order n if its nth order statistics are almost periodic functions of time. The inherent periodicity can be quantified by Fourier analysis of the relevant functions listed in Table 2. Cyclic autocorrelation function (CAF) and spectral correlation density (SCD) are commonly applied to identify second-order (or wide-sense) cyclostationary stochastic processes whose autocorrelation functions vary periodically with time [84]. In addition, they are the foundation for kernel-based cyclostationary introduced in Section V-F. The definition of autocorrelation function is given by (26), and its Fourier coefficients make up the CAF shown in (27).

$$R_{xx}[\tau] = \sum_{-\infty}^{\infty} x[n]x[n-\tau] \tag{26}$$

$$R_x^{\alpha}[\tau] = \lim_{T\to\infty} \frac{1}{T} \sum_{n=-T/2}^{T/2} x[n]x[n-\tau]e^{-j2\pi\alpha n} \tag{27}$$

here α is the cyclic frequency. This can be viewed as extraction of cycle frequencies from the autocorrelation function $x[n]x[n-\tau]$. Each α corresponds to a particular Fourier coefficient, which means that α gives the frequency at which the autocorrelation function contains some power. These values are derived from the sine waves generated by the products of two first-order cyclostationary functions.

The SCD is the FT of CAF along the τ-axis, defined as:

$$S_x^{\alpha}[f] = \sum_{\tau} R_x^{\alpha}[\tau]e^{-j2\pi f\tau} \tag{28}$$

here f is the spectral frequency. SCD allows locating the amount of time correlation between frequency-shifted versions of discrete signals in the frequency domain. Note that the range of support for the SCD is restricted to the region presented in Fig. 3, and Fig. 4 presents the distribution differences in different modulations' SCD plots.

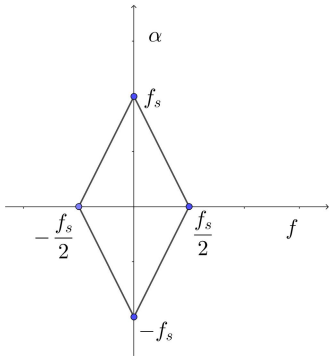


FIGURE 3. Region of support for SCD.

The 2D representation of the SCD shown in Fig. 3 has symmetry and redundancy [85]. It can be collapsed along one of the axes to give the 1D f-profile or α-profile [86], that is:

$$I(\alpha) = \max_f |S_x^{\alpha}[f]| \quad I(f) = \max_{\alpha} |S_x^{\alpha}[f]| \tag{29}$$

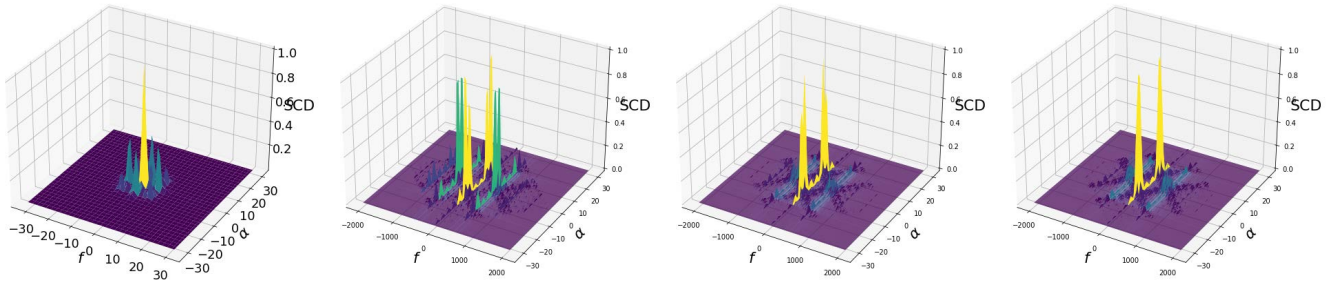
B. SECOND ORDER CYCLOSTATIONARY FEATURES FOR AMC

Since the computational complexity of traditional spectral correlation is high, there are many existing techniques designed for simpler algorithms, such as FFT Accumulation Method (FAM) [87], Strip Spectral Correlation Algorithm (SSCA) [88], and Faster Spectral Correlation (FSC) [89]. However, although SCD can now be implemented algorithmically or in hardware relatively speedily, its limitation is also apparent. For modulations whose amplitudes or phases satisfy the symmetry of the constellation diagram, based on the properties of prototype functions, their SCD peaks appear at the same positions, which makes SCD not discriminative for modulations such as M-PSK and M-QAM ($M \geqslant 4$), as their SCD plots have the same formats as QPSK [90], [91]. Fig. 4 also illustrates this point, for QPSK, higher order M-QAMs and APSKs, SCD does not have excellent discrimination ability.

But for FSK and ASK, SCD still has good capabilities. In [92], a real value matrix is obtained by taking the magnitude of the SCD output. Then, using the 2D FFT and inverse FT, the SCD image is mapped onto a 64 $\times$ 64 image. Due to symmetry and sparse properties of the SCD, only select a triangular region which is one-eighth of the scaled image size is required. Next, the 64 $\times$ 8 pixel data is reshaped into a vector with 512 lengths. Lastly, the vector is normalized to zero mean and unit variance, before being put into the classifier network. The modulations tested were 4FSK, 16QAM, BPSK, QPSK, and OFDM, a set which is not as challenging as some others. Accuracy achieved was already 90% for an SNR of -2dB. In [93], α-profile is used as the input features for a neural network classifier. The modulation set contains 2FSK, 4FSK, 8FSK, BPSK, QPSK, MSK, and 2ASK. When SNR is higher than -5dB the recognition rate of 2FSK, 8FSK, QPSK, 2ASK is more than 95%. The overall accuracy is more than 90% when SNR is higher than 0dB. It is evident that the application of SCD can help with certain types of modulation classification.

C. HIGH ORDER: CYCLIC MOMENTS AND CUMULANTS

One important issue when utilizing scalar HOS is that important information is lost by trying to summarise a complex waveform with a few numbers. The HOS values will have large data-dependent fluctuations making them inappropriate for classification. Hence scalar HOS are not well suited for discriminating between different circular constellation modulation formats [18]. In this case, high order cyclostationary properties can be useful. The nth order



(a) OOK (b) BPSK (c) QPSK (d) M-QAM

FIGURE 4. SCD plots for different modulation types.

lag-product is the nonlinear transformation in (3), defined as:

$$L[n,\underline{\tau}]_m = x^{(*)}[n+\tau_1]x^{(*)}[n+\tau_2]\ldots x^{(*)}[n+\tau_m] \quad (30)$$

where $\underline{\tau}$ denotes individual delays and the notation $x^{(*)}[n]$ means optional complex conjugation.

Suppose a complex signal is nth order periodic in frequency α. In that case, the averaging of the lag-product, multiplied by a complex exponential of this frequency conducts to the definition of cyclic temporal moment function (CTMF), which is expressed as:

$$R_x^{\alpha}(\underline{\tau})_m = \lim_{T\to\infty}\frac{1}{T}\sum_{n=-T/2}^{T/2} L[n,\underline{\tau}]_m e^{-j2\pi\alpha n} \quad (31)$$

this can be seen as discrete Fourier transform of $L[n,\underline{\tau}]_m$, particularly when $m = 2$ the CTMF is equivalent to CAF. At the same time, according to (23), the cyclic temporal cumulant function (CTCF) can be obtained through the CTMF (31). Take a zero-mean random signal as an example, because M_{20} is the mean of a sequence $M_{20} = 0$. In this case C_{40} is approximately equivalent to its M_{40}.

D. RD-CTCF

A wide range of studies set $\underline{\tau}$ in (30) to small values or zero for several reasons [10], [79], [80]: the $C_x^{\alpha}(\underline{\tau})_n$ is not absolutely integrable for all $\underline{\tau}$; with increasing τ, the value of the high order cumulants tends to be close zero; and the difficulty of implementation. In the light of these, the CTCF with $\underline{\tau} = 0$, is termed reduced dimension cyclic temporal cumulant function (RD-CTCF).

1) FOURTH ORDER CYCLIC FUNCTION

Depending on (23) and (31), the simplified fourth order cyclic cumulants can be calculated through:

$$\begin{aligned} C_{40}^{\alpha}(\underline{\tau}=0) &= \frac{1}{N}\sum_n s[n]^4 e^{-j2\pi n\alpha} \\ &\quad - 3\left(\frac{1}{N}\sum_n s[n]^2 e^{-j2\pi n\alpha}\right)^2 \\ C_{42}^{\alpha}(\underline{\tau}=0) &= \frac{1}{N}\sum_n s[n]^2 s^*[n]^2 e^{-j2\pi n\alpha} \\ &\quad - \left|\frac{1}{N}\sum_n s[n]^2 e^{-j2\pi n\alpha}\right|^2 \\ &\quad - 2\left(\frac{1}{N}\sum_n s[n]s^*[n]e^{-j2\pi n\alpha}\right)^2 \end{aligned} \quad (32)\ (33)$$

The plots of $C_{40}^{\alpha}(\underline{\tau}=0)$ and $C_{42}^{\alpha}(\underline{\tau}=0)$ are presented in Fig. 5. The x-axis is cyclic frequency and y-axis is magnitude of complex cumulants.

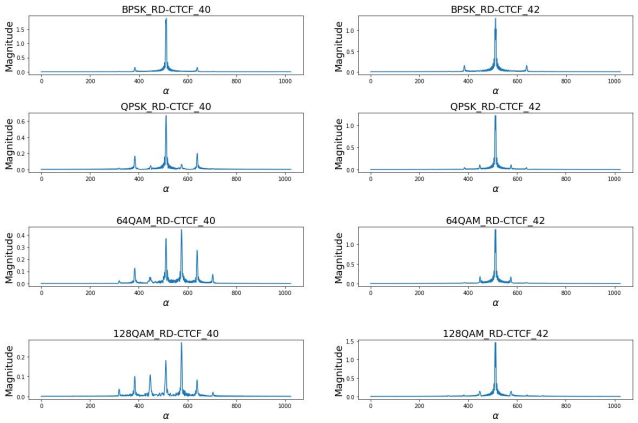


FIGURE 5. Fourth order RD-CTCF using (32) and (33).

2) SIXTH ORDER CYCLIC FUNCTION

Similar to fourth order, the sixth order cyclic functions are derived from (23) and (31). Again, the plots of $C_{61}^{\alpha}(\underline{\tau}=0)$ and $C_{63}^{\alpha}(\underline{\tau}=0)$ are shown in Fig. 6. The x-axis is cyclic frequency and y-axis is magnitude of complex cumulants.

3) WINDOW SIZE AND OVERLAP

The above implementations all apply sliding windows as described by the following equation:

$$m_{pq}^{\alpha}(\underline{\tau}=0) = \frac{1}{N_w}\frac{1}{N}\sum_w\sum_n w[s[n]^{p-q}s^*[n]^q]e^{-j2\pi n\alpha} \quad (34)$$

As mentioned before, the window size and overlap should be adjusted according to the specific situation. When the

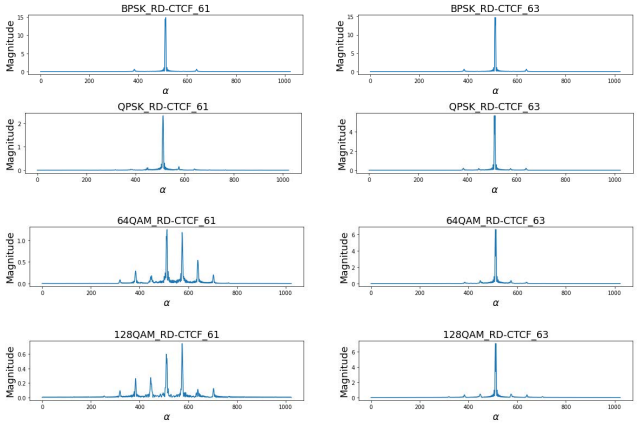


FIGURE 6. Sixth order RD-CTCF.

window size and the total length of a signal are fixed, the larger the overlap length, the greater the time resolution [46].

E. CTCF FEATURES FOR AMC

The previous subsection showed that the CTCF extracts important features, and by observing the number, amplitude, and position of its spectral lines, signal detection and classification can be carried out. Most CTCF applications have been fourth- and sixth-order, and the signal can be detected and classified by examining its spectral line number, peak magnitudes and peak locations. Table 5 presents a summary of CTCF-based approaches, but for the current results, this operation is not yet widely applied in AMC.

F. KERNEL-BASED

Second-order cyclostationary approaches have been used extensively, especially for spectral sensing and automatic modulation classification tasks [95]. Nevertheless, the second-order cyclostationary (SCD or CAF) properties are not suitable for non-Gaussian communication channels, resulting in the deterioration performance, e.g. AMC of QAM in time-varying channels [96]. High order cyclostationary features can be further extended to kernel-based ones which can subsume both second and higher order features in a computationally efficient manner [97].

1) CORRENTROPY

Correntropy as a method called information theoretic learning (ITL) was first applied to machine learning problems by [98] and was found to have very different properties compared to mean square error (MSE), which is very useful in nonlinear, non-Gaussian signal processing. Similar to CAF which measures similarity across lags as the autocorrelation, correntropy yields the entropy of the random variable during averaging of the lags. Thus, the correntropy covers the higher order moments of the probability density function (PDF) but it is much simpler to estimate directly from the data and bypasses the need for traditional moment expansion [95].

Correntropy is a similarity measure between two arbitrary random variables X and Y: $V_\sigma(X, Y) = E[\kappa_\sigma(X - Y)]$ where $\kappa_\sigma(\cdot)$ is usually assumed to be a normalized Gaussian kernel with variance σ.

$$V_\sigma(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} E\left[\exp\left(\frac{-(X-Y)^2}{2\sigma^2}\right)\right] \tag{35}$$

For applications in signal processing, let $X = x[n]$ and $Y = x[n + \tau]$. Applying the Taylor series expansion of the exponential function in the Gaussian kernel, we get:

$$V_\sigma[n, \tau] = \frac{1}{\sqrt{2\pi}\sigma} \sum_{m=0}^{\infty} \frac{(-1)^m}{2^m \sigma^{2m} m!} E[(x[n] - x[n+\tau])^{2m}] \tag{36}$$

σ is also denoted as Gaussian kernel size. Additionally, if the correntropy function $V_\sigma[n, \tau]$ is periodic with T_0, then the Fourier series expansion can be applied:

$$V_\sigma[n, \tau] = \sum_\alpha V_\sigma^\alpha[\tau] e^{j2\pi\alpha n} \tag{37}$$

$$V_\sigma^\alpha[\tau] = \frac{1}{T_0} \sum_n V_\sigma[n, \tau] e^{-j2\pi\alpha n} \tag{38}$$

here $\alpha = k/T_0$ is the cyclic frequency that represents all multiples of fundamental frequency, where $k \in \mathbb{Z}$. The Fourier coefficient $V_\sigma^\alpha[\tau]$ of correntropy function is called cyclic correntropy function (CCF). From τ and α, the spectral function of CCF can be generated through the Fourier transform, this is defined as cyclic correntropy spectral density (CCSD):

$$S_x^\alpha[f] = \sum_\tau V_\sigma^\alpha[\tau] e^{-j2\pi f\tau} \tag{39}$$

It can be seen in (36) that the Gaussian kernel is sensitive to higher-order statistic information. Thus an appropriate value for σ must be selected for use in the correntropy kernel [31]. Fig. 7 shows CCSD plots for $\sigma = 0.3$. It can be seen that the CCSD plots are able to better discriminate modulations types than SCD (Fig 4), and the high order components in the Gaussian kernel help distinguish between M-QAM types.

2) HYPERBOLIC TANGENT

Hyperbolic Tangent Correlation (HTC) is a new kernel function method, proposed in [99]. This technique has been used mainly for denoising and we are not aware of its applications in AMC. Its expression is similar to original correntropy:

$$T[n, \tau] = E[\tanh(x[n]x[n+\tau])] \tag{40}$$

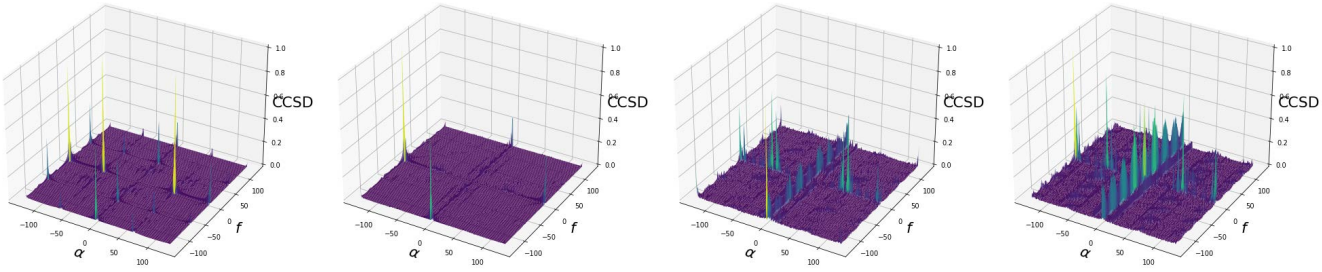
Its cyclic correlation function is called hyperbolic-tangent cyclic correlation (HTCC), and the spectrum of the HTCC is:

$$H_x^\xi[\tau] = \frac{1}{T_0} \sum_n T[n, \tau] e^{-j2\pi\xi n} \tag{41}$$

$$Z_x^\xi[f] = \sum_\infty H_x^\xi[\tau] e^{-j2\pi f\tau} \tag{42}$$

TABLE 5. Summary of key cyclic high-order features used.

Citation	Modulation Set	Key Features	Main Task	Evaluation
[18]	BPSK, MSK, π/4-QPSK, QPSK, OQPSK, 16-QAM	RD-CTCF$_m$ (m=20,21,40,42)	AMC	Proposed a hierarchical classifier robust to synchronization error, phase errors, and flat fading channel.
[80]	4-ASK, 8-ASK, BPSK, QPSK, 8-PSK, 16-PSK, 16-QAM, 32-QAM, 64-QAM	RD-CTCF$_{8X}$ (X=0,2,4,6,8)	AMC	Robust to single- and multi-antenna in flat fading channels. But large number of samples are required.
[21]	4-QAM, 16-QAM	RD-CTCF$_{42}$, RD-CTCF$_{63}$	QAM Identification	Steady with respect to timing, phase, frequency offsets, and phase noise, but the QAM's classes are too few.
[79]	PAM, QAMs, QPSK	Fourth-order cyclostationary	Spectra Sensing	Shows great anti-noise ability but too few modulation types.
[94]	Sub-carriers of OFDM	RD-CTCF$_{40}$	Identification of the OFDM sub-carriers	Use four-order cumulants to accurately identify the number of sub-carriers under various OFDM-based elastic optical networking.



(a) BPSK (b) QPSK (c) 8QAM (d) 64QAM

FIGURE 7. CCSD for different modulation types ($\sigma = 0.3$).

The negative exponential in the Gaussian kernel can only map real sequences to positive values, while the tanh function includes negative ones. This matches the correlation function, benefiting signal separation and its mathematical derivation [99]. Moreover, HTCC requires no parameters while correntropy requires setting the parameter σ.

G. KERNEL-BASED FEATURES FOR AMC

The main research on applying cyclic correntropy to AMC has been by Fontes *et al.* [95], [97] and Câmara *et al.* [100]. They showed mathematically that CCF contains more information other than CAF [97], including high order information, and a sinusoidal function which controls the phase shift of the response regardless of the type of random process. These properties serve to make the cyclic correntropy relatively insensitive to non-Gaussian noise, e.g. impulsive noise.

In [95], the authors study the correntropy function itself instead of CCF or CCSD. Due to Gaussian kernel's non-linear transformation, cross entropy is not guaranteed to have zero mean. Gaussian kernel based cross entropy coefficients [101] between the input signal and pre-stored known modulation signals are utilized to circumvent this. Also, because the σ value controls the behaviour of the kernel function, using a classifier based on the cross entropy coefficients, the authors investigated the relationship between SNR and kernel size for MSK, OOK, BPSK, QPSK and 16-QAM. For better performance, the most suitable kernel size can be chosen according to the circumstance, as shown in Table 6.

TABLE 6. Relationship between SNR and cross-entropy coefficients' kernel size σ [95].

SNR(dB)	MSK	OOK	BPSK	QPSK	16-QAM
-10	4.64	4.64	46.41	10	21.54
-5	2.21	4.64	2.15	10	10
0	0.21	1	1	0.46	2.15
5	0.1	0.21	0.46	0.1	0.1
10	0.1	0.96	0.1	0.1	0.1
15	0.1	2.15	0.1	0.1	0.1

In addition to these coefficients, CCF and CCSD also have a wide range of characteristics to analyze. For example, cyclic components at specific α-axis positions can be employed as signatures [100]. Commonly, the most reliable method is still to directly recognize the entire 2-dimension CCSD plot (with spectra frequency f vs. cyclic frequency α) as a feature map. A summary of researches developed using cyclic correntropy are presented in Table 7. In fact, even simple classifiers have good accuracy (close or equal to 100%) at high SNR, so the major challenge in employing correntropy is to make the feature map robust at low SNR. Collectively, CCF and CCSD are mostly not affected by impulsive noise and provides good discrimination of different modulation schemes.

TABLE 7. Summary of kernel-based cyclostationary techniques for AMC.

Citation	Modulation Set	Features	Classifier	Evaluation
[102]	2ASK, BPSK, QPSK and 16QAM	• f-profile: Select the maximum of $\|CCSD\|$ among α-axis • Specific profiles: Extract $\|CCSD(f)\|$ at $\alpha = 0, fc, \pm 2fc, 4fc$	Firstly using PCA to re-extract the key features from these f-profiles, and then employ multi-layer RBF to classify.	Several samplings are performed among α-axis of CCSD to obtain diverse f-profiles. A lightweight classifier is applied instead of a deep model. The outcome has good performance against impulsive noise. The loss of accuracy is minimized with varying characteristic exponents. But the considered modulation types are fewer.
[30]	BPSK, 2ASK, QPAK, 8ASK and 16QAM	Directly employ the 2D-CCSD as graph.	ResNet-based model [103].	From previous work [102], utilize CCSD plots and deep ResNet for solving non-Gaussian noise problem.
[100]	BPSK, QPSK, 8QAM, 16QAM and 32QAM	$\|CCF\|$ values at $4f_c$ combined with correlation coefficients.	Decision Tree-like classifier	It has high accuracy for the modulations other than 16QAM, with BPSK, 8QAM and 32QAM being nearly 100% at SNR -2dB. But there is still room for improvement for 16QAM and QPSK. Also demonstrated that the optimal value of kernel size is 1.2 for CCF-based classifier.
[31]	BPSK, QPSK, 2FSK, 4FSK, MSK, 16QAM, 64QAM, 8PSK, 2ASK and 4ASK	• 2D-CCSD M*P array (M channels: α ranges; P length: f ranges) • 1D profile: Select the maximum of $\|CCSD\|$ among f-axis	A customized NN named LSMD: Combination of LSTM, convolution layers, dense layers, pooling layers and batch normalization.	Elaborate narrative, effective combination of existing NN modules and correntropy algorithms. It takes full advantage of the anti-noise ability of correntropy. Achieves good accuracy among all SNR ranges.

VI. SUMMARY OF PUBLISHED RESULTS

This section reviews the published techniques included in the taxonomy in Table 1. The overall comparison is summarised in Table 8, and lists the configurations and final accuracy of a number of key papers on the subject. Note that in the SNR & Accuracy column, the accuracy is from the lowest to the last stable value; the corresponding SNR ranges do not include the SNRs after the accuracy becomes stable. This is employed to compare the performances better.

A. FUSION OF FEATURES

In addition to directly employing the aforementioned features, numerous studies applied combinations of features.

1) FUSION OF SCALAR FEATURES

Fused scalar features often use statistics such as standard deviation of the absolute value of normalized centered instantaneous amplitude, or the value of specific locations of some transformations, such as CWT and CTCF. In [66], some statistics of CWT and HOC are used as nodes thresholds in a decision tree. Lee *et al.* [109] directly combines HOC and signal statistics into 28 numerical values and feeds them into dense layers to classify. Majhi *et al.* [18] extracts the peak positions of RD-CTCF$_{20}$, RD-CTCF$_{21}$, RD-CTCF$_{40}$, RD-CTCF$_{42}$, and merges them with HOC as features.

2) FUSION OF MULTI-DIMENSION FEATURES

The flexibility of neural networks enable them to readily handle data of different dimensions. For example, [48] uses separate CNNs for feature extraction of raw IQ and FFT, and then employs a fully connected layer as a fusion layer to combine and map these CNN outputs to different categories. Wang *et al.* [116] combines IQ samples and IQ-converted constellation graphs as inputs to a CNN, and confirms that a combination of features can have better accuracy by showing it outperforming IQ sequences. On the other hand, [117] shows that for raw IQ inputs, a CNN and LSTM will have similar performance, and the application of HOC can improve the accuracy by around 8%. Inspired by the former's research, [118] also utilizes IQ sequence and HOC as inputs to a DNN classifier and achieves a higher accuracy but has limited accuracy for M-QAM classification. For combinations of sequences, the discrete orthonormal Stockwell transform [104] and the IQ series are integrated as two channels and fed into a CNN; instead of using two separate NN modules ahead of a fusion module; this uses fewer parameters. For the fusion of 2D images, the performances of using smooth pseudo Wigner-Ville distribution (SPWVD) [119] and Born-Jordan distribution (BJD) alone are found to be similar [107]. They use two individual CNN to process each feature image and connect these two CNN's outputs with fully-connected layers.

Thus mixed features are often more effective than using them separately. Unfortunately, such models inherently require more parameters than a single model.

B. OBSERVATIONS

As can be seen in Table 8, the SNR ranges, modulation sets, and classifiers of different studies make direct comparison impossible. Hence for the sake of standardization,

TABLE 8. Performance summary of pre-processing based AMC algorithms.

Cita-tion	Features	Modulations	Classifier	Channel	SNR(dB) and Accuracy	Additional Information
[48]	Raw IQ and FFT	QPSK, 8PSK, QAM16, QAM64, WBFM, AMDSB, CPFSK, GFSK	Customized CNN	Fading and AWGN	[-20, +18] 11% - 86.97%	Platform: Tensorflow; Execution time: 0.2351ms per sample
[104]	Raw IQ and discrete orthonormal Stockwell transform	BPSK, QPSK, 8PSK, GMSK, 16APSK, 64APSK, 16QAM, 64QAM, AM-DSB-WC, FM	Multi-layers convolution neural network	Fading and AWGN	[-8, +8] 50%-97.3%	Platform: Keras & Google Cloud VM
[105]	Cyclic spectra and constellation as images	8PSK, AM-DSB, AM-SSB, BPSK, CPFSK, GFSK, PAM4, QAM16, QAM64, QPSK, WBFM	Multi-layers convolution neural network	AWGN	[-20, +10] 16%-92%	Platform: NVIDIA GeForce GTX 1060 GPU
[59]	Peak counts of CWT	M-QAM (M=16,64,128)	Thresholds	AWGN	[-4, 0] 93.1% - 100%	Platform: Matlab
		M-ASK (M=2,4,8)			[-10, -2] 91.5% - 100%	
[60]	Variance and kurtosis of CWT	QAM, PSK, FSK	Thresholds	AWGN	[0, +15] 64% - 100%	Platform: Matlab
[64]	CWT sequences	BASK, BPSK, BFSK	Similarity score compared with existing templates	AWGN	[-10, -1] 97.34% - 100%	Platform: Matlab
[65]	Local maximum of CWT and other statistics	2ASK, 4ASK, 8ASK, 16QAM, 32QAM, 64QAM, BFSK, QFSK, 8FSK, BPSK, QPSK and 8PSK	Decision tree	AWGN	[-5, +15] 73.87% - 100%	Platform: Matlab
[66]	Statistics of CWT and HOC	2ASK, 4ASK, 8ASK, 2FSK, 4FSK, 8FSK, 2PSK, 4PSK, 8PSK and 16QAM	Decision tree	AWGN	[0, +10] 56.7% - 100%	Platform: Matlab
[70]	Energy and entropy from decomposed coefficients of S-transforms	BPSK, QPSK, FSK, MSK	NN, Naive Bayes, LDA, SVM, k-NN	AWGN	[0, +20] 30%-98%	Execution time: 8.07ms per sample
[69]	Correlation coefficients and variances of S-transform	CP, LFM, BPSK, BFSK	Hierarchical tree-like	AWGN	[+3, +10] 97%-100%	\
[106]	Time-frequency map and instantaneous autocorrelation map	LFM, SF, BPSK, QPSK, 2FSK, 4FSK	Multi-layers convolution neural network	AWGN	[+6, +14] 97%-99.8%	Platform: Matlab
[107]	Smooth pseudo wigner-ville distribution and Born-Jordan distribution	BPSK, QPSK, 2FSK, 4FSK, 2ASK, 4ASK 16QAM, OFDM	ResNet-152	AWGN	[-4, +8] 92.2%-98.5%	Platform: Matlab & Python3.5
[71]	Energy entropy of S-transform	16QAM, QPSK, 4FSK, OFDM, FM, SSB	SVM	AWGN	[- 10, +5] 46%-100%	Platform: Matlab
				Rayleigh fading	[0, +30] 65%-97.5%	
[108]	Eye diagrams as images	4PAM, RZ-DPSK, NRZ-OOK, RZ-OOK	Multi-layers convolution neural network	AWGN	[+10, +25] 59%-100%	\

TABLE 8. *(Continued.)* **Performance summary of pre-processing based AMC algorithms.**

[92]	Flatten specific area of SCD to a vector	4FSK, 16QAM, BPSK, QPSK and OFDM	Deep Belief Network (DBN)	AWGN	[-2, +5] 90% - 99.6%	Platform: Matlab
[93]	α-profile	2FSK, 4FSK, 8FSK, BPSK, QPSK, MSK and 2ASK	Multi-layer perception	AWGN	[-10, +5] 54.86% - 95%	Platform: Matlab
[102]	α-profile and some specific frequency profiles	2ASK, BPSK, QPSK and 16QAM	PCA and RBF	Impulsive noise	[-5, +5] 37% - 100%	\
[30]	CCSD plots	BPSK, 2ASK, QPAK, 8ASK and 16QAM	ResNet-based	Impulsive noise	[-5, +5] 37.5% - 100%	Platform: Matlab and Tensorflow; NVIDIA GTX 1080 Ti GPU
[31]	CCSD plots essentially	BPSK, QPSK, 2FSK, 4FSK, MSK, 16QAM, 64QAM, 8PSK, 2ASK and 4ASK	Customized NN named LSMD: LSTM, convolution layers, dense layers, pooling layers and batch normalization	Fading and AWGN	[-20, +2] 28% - 100%	13.97ms per sample on Nvidia GeForce GTX 1080 GPU
[100]	Correlation coefficients of CCF and CCF value at specific positions	BPSK, QPSK, 8QAM, 16QAM and 32QAM	Decision Tree-like	Impulsive noise	[-10, +6] 0% - 100%	\
[109]	HOC and signal statistics 28 in total	BPSK, QPSK, 8PSK, 16QAM and 64QAM	Dense NN based	AWGN and Rayleigh fading	+5 only 86.43%	\
[110]	Fourth order cumulants	BPSK, 8PSK, 4QAM, 16QAM, 64QAM	DNN based	Multi-fading channels	[0, +15] 83% - 100%	Platform: Matlab
[111]	HOC, 6 in total	ASK, 4ASK, 2FSK, 4FSK, 2PSK and 4PSK	DNN based	AWGN and flat fading	[-5, -2] 99.5% - 100%	\
[112]	HOC, 5 in total	BPSK, QPSK, 8PSK, 16QAM and 64QAM	DNN based	AWGN	[-5, +10] 85.61% - 99.7%	\
[113]	HOC, 12 in total	BPSK, QPSK, 8PSK, 16QAM, 64QAM and 256QAM	DNN based	Fading and AWGN	[0, +12] 87% - 100%	Platform: Matlab
[114]	C_{42} and C_{80}	2PSK, 4PSK, 8PSK, 16PSK, 4QAM, 16QAM, 64QAM	DNN based	Multi-path fading channels	[0, +20] 57.15% - 76.58%	\
[115]	HOC, 7 in total	BPSK, QPSK, 16QAM, 64QAM	KNN	AWGN	[0, +15] 71% - 100%	Platform: Matlab/GPLab
[12]	C_{42} and C_{63}	BPSK, QPSK, 8PSK, 16QAM, 64QAM and 256QAM	Polynomial classifier	AWGN	[0, +20] 56% - 98%	\
[18]	RD-CTCF$_m$ m=20,21,40,42 peak positions and element HOC	BPSK, MSK, $\pi/4$-QPSK, QPSK, OQPSK, 16QAM	Hierarchical tree-like	Flat fading channel	[0, +5] 16.7% - 100%	Platform: NI PXIe; Algorithm complexity: $O(NlogN)$
[80]	RD-CTCF$_{8X}$ X=0,2,4,6,8 at specific α values	4ASK, 8ASK, BPSK, QPSK, 8PSK, 16PSK, 16QAM, 32QAM, 64QAM	Probability score using existing templates	AWGN and flat fading channels	[0, +15] 40% - 95%	\
[21]	RD-CTCF$_{42,63}$ at specific α values	4QAM and 16QAM	Probability score using existing templates	Phase noise and fading channel	[-5, +5] 60% - 100%	\
[73]	C_{20}, C_{40}, C_{42}	BPSK, PAM(4), PSK(8), 16-QAM	Hierarchical tree-like	AWGN and impulsive noise	[-5, +10] 25%-100%	\

in following section, we reproduce the pre-processing techniques and compare them using a unified classifier. Nevertheless, we have formed the opinion that: 2D features outperform 1D and scalar ones; combined features offer improvements over uncombined ones; second-order transforms are more robust to noise than linear ones; and second- and high-order techniques perform better than first-order.

VII. EXPERIMENTS: REPRODUCTION AND OBJECTIVE COMPARISON

A. EXPERIMENTAL CONFIGURATION AND PYTHON LIBRARY

In addition to the above works, we also provide a systematic open-source Python library for computing the aforementioned pre-processing algorithms, designed to allow access to specific functions or to enable the execution of all functions available in the library. It is also possible to customize the class methods and attributes to introduce new functions for future research works or practical applications. This library focuses on the second and higher order cyclostationary features in Table 1. For other transformations we directly call some off-the-shelf libraries. The outputs of different pre-processing algorithms can be fed into an LDA classifier to calculate the recognition accuracy, which is helpful for the comparison of the accuracy of different feature maps. This library is available from: https://github.com/louisinhit/AMC_preprocess.

Since the datasets, SNR ranges and classifiers used in previous studies are all different, we conducted a comparison using our own library and a unified dataset. I&Q input signals

TABLE 9. Information on modulation data set.

Modulation Classes	OOK, 4ASK, 8ASK, BPSK, QPSK, 8PSK, 16PSK, 32PSK, 16APSK, 32APSK, 64APSK, 128APSK, 16QAM, 32QAM, 64QAM, 128QAM, 256QAM, GMSK, OQPSK for digital.
3 SNR Group	Below –5 dB; -5 to + 10 dB; and above +10 dB.
Number of Samples	19 classes, 26 SNR slots, 200 samples per SNR per modulation, totally has $19 * 26 * 200$ signal slices.
Raw Data Size	Each slice has 1024 sample points in IQ format with shape [2, 1024].

were uniformly selected from the 19 digital modulations from DeepSig RadioML 2018.01A dataset [25] to produce a benchmark set. We divide the SNR range into 3 groups, low, middle and high. Details are given in Table 9. The split between the train set and the test set is 9:1. LDA is chosen here as the classifier [120] because there are no parameters involved, and a linear technique leads to a straightforward and simple comparison.

B. EVALUATION OF HIGH ORDER STATISTICS

For determining the optimal normalization approach, in the baseline model, we apply general normalization (43) to 14 HOS: $\{|C_{20}|, |C_{21}|, |C_{40}|, |C_{41}|, |C_{42}|, |C_{60}|, |C_{61}|, |C_{62}|, |C_{63}|, |C_{80}|, |C_{82}|, |C_{84}|, |M_{101}|, |M_{103}|\}$.

$$\left|\frac{C_{pq}}{C_{21}^{\frac{p}{2}}}\right| \tag{43}$$

The HOS processing approaches mentioned in other papers and the accuracy they achieved using the LDA classifier are shown in Table 10. It is clear that the n-th root of magnitude value $|C_{pq}|^{\frac{2}{p}}$ outperforms the others, therefore, this method should be used for scalar HOS.

TABLE 10. Experiment: LDA accuracy for element HOS.

Group	Normalization Approaches	Overall Accuracy
Baseline	14 HOS using general normalization	53.8%
1 (from [14])	Substitute $\|C_{40}\|$ with $\|C_{40}\|/\|C_{42}\|$ Substitute $\|C_{63}\|$ with $\|C_{63}\|^2/\|C_{42}\|^3$	50.74%
2 (from [121])	Substitute $\|C_{63}\|$ with $\|C_{63}^2/C_{42}^3\|$	52.47%
3 (from [12])	All use n-th root i.e. $\|C_{pq}\|^{\frac{2}{p}}$	**57.36%**

C. EVALUATION OF TIME-FREQUENCY TRANSFORMATIONS

1) OVERALL COMPLEXITY COMPARISON

Finally we carefully compare the transformations introduced in this paper. Details of the signal characteristics adopted in this comparison experiment are shown in Table 11. It provides the theoretical computation complexity and the time required to process a 1024 long sample points segment on Intel(R) Xeon(R) CPU E5-2630 v4 platform. Here the complexity of the add-window operation is skipped and the focus is on the complexity involved in the mathematical operation.

In order to control the variables as much as possible for objective and reliable results, all performance test experiments are reproduced locally. The modulation set is shown in Table 9. The parameters for each technique are selected to be as close to the suggested values in the literature, e.g., in scalar HOS, the best normalization algorithm is chosen; in DWT or CWT, the most appropriate scale factor is used; the fast CCSD algorithm follows [97], and adopts the optimal σ value; and fast SCD is calculated according to [87]. According to Table 11, a higher order transformation does not mean more computational overheads. The high complexity of the cyclic-correlation spectra stems mainly from the two Fourier transforms for the t and τ axes, respectively, or the computation of the corresponding kernel functions.

2) INFLUENCE OF DIFFERENT SNR RANGES

Fig. 8 shows the overall experiment results for different approaches. The horizontal axis labels correspond to the pre-processing techniques and the size of the features. Different colors denote the three SNR groups. Here, Raw IQ represents the accuracy for recognizing the original signals

TABLE 11. Details of the features in experiment.

Label	Annotation	Feature Size	Theoretical Computation Complexity	Execution Time
Raw IQ	Original signal	(1024, 2)	–	–
FT	Fourier transform	(1024)	$NlogN$	0.000306s*
Level1 DWT	Level1 Haar Discrete WT	(512, 2)	N	0.000289s†
CWT	Continuous WT	(1024, 4)	N	0.001978s†
STFT	Short time Fourier transform	(256, 9)	$PN'logN'$	0.001799s*
SCD profiles	SCD with multi α- and f- profiles	(512, 8)	$PN'logN' + 2PN' + N'N'P + \frac{1}{2}N'N'PlogP$ [87]	0.047724s
SCD graph	SCD array as images	(512, 512)		
CCSD profiles	CCSD with multi α- and f- profiles	(512, 8)	$3PN'N' + 2N'N'logN' + 4N'N'$ [97]	0.020375s
CCSD graph	CCSD array as images	(512, 512)		
HTC profiles	Hyperbolic tangent with α- and f- profiles	(512, 8)	$3PN'N' + 2N'N'logN' + 4N'N'$ [99]	0.027078s
HTC graph	Hyperbolic tangent as images	(512, 512)		
Element HOS	14 singular high order statistics	(14)	MN [18]	0.001971s
RD-CTCF	Four different orders of RD-CTCF	(1024, 4)	$MNlogN$ [18]	0.047792s

Here, N is the total signal length, N' is window size, P is the number of sliding windows when involving in windowing, and M is the order of the high order transform. For the execution time, * means applying the Scipy library [122], † means applying the PyWavelets library [123], and the others are just built from scratch using Numpy [124]. Direct calls to existing libraries can greatly improve efficiency, so it's necessary to refer to both Theoretical Complexity and Execution Time. Note the complexity of adding-window operation is skipped here, the focus is on the mathematical operation.

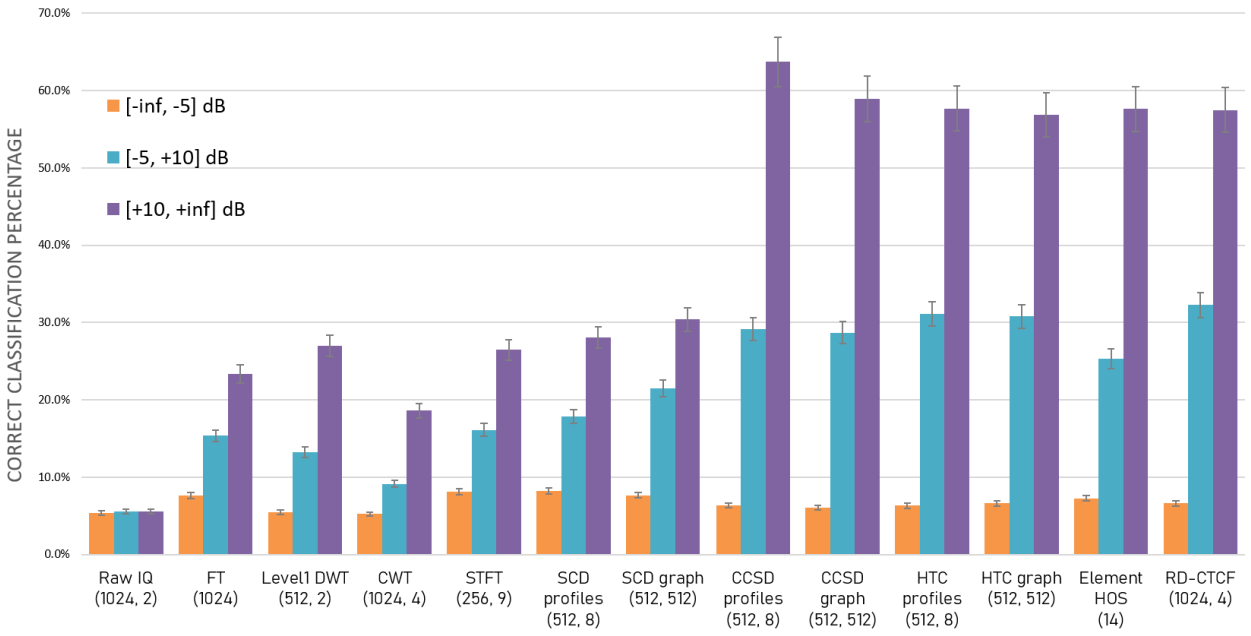
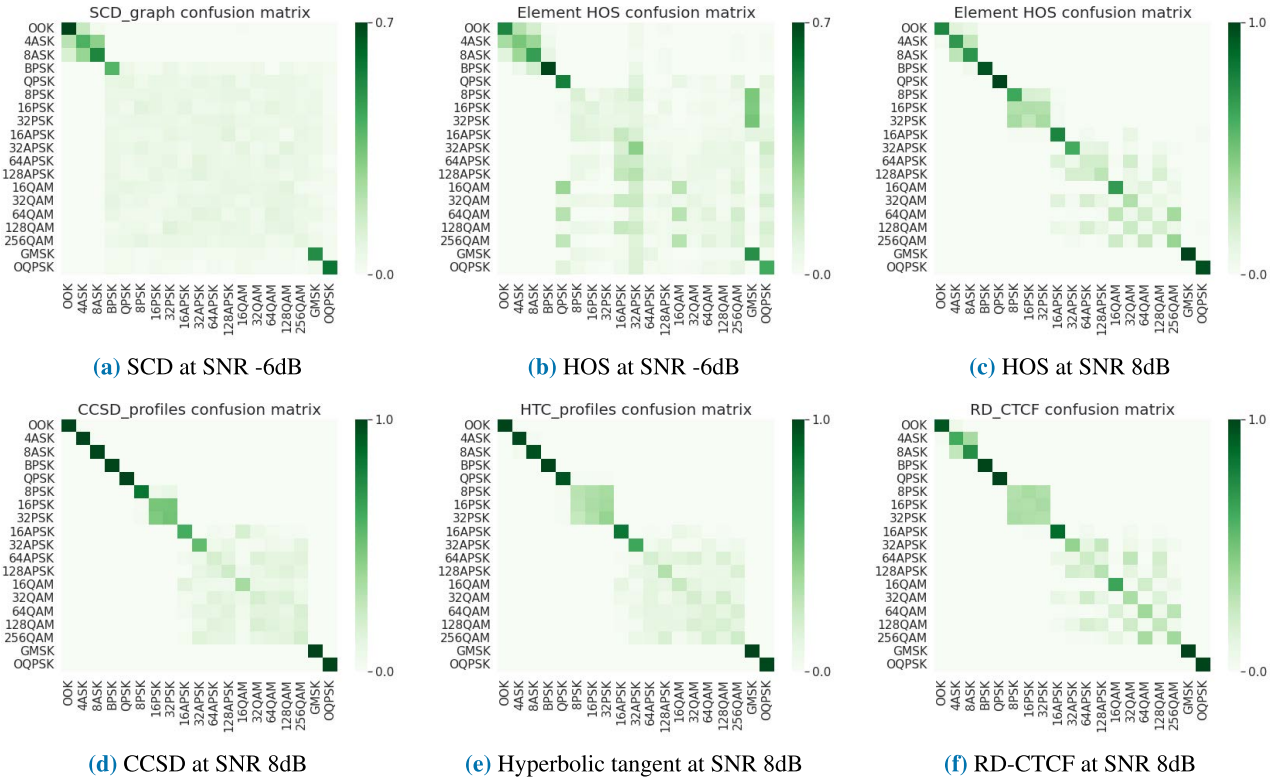


FIGURE 8. Accuracy comparison of different pre-processing techniques using one-layer LDA classifier. Different colors correspond to different SNR ranges. Features details are in Table 11.

without any pre-processing, which serves as a baseline. Most methods are able to improve over the baseline, with cyclostationary analysis and the higher-order transforms being generally better than those of the linear time-frequency transform. Interestingly, the advantage of pre-processing is greatly weakened at low SNR (below -5 dB). The CCSD,



(a) SCD at SNR -6dB (b) HOS at SNR -6dB (c) HOS at SNR 8dB

(d) CCSD at SNR 8dB (e) Hyperbolic tangent at SNR 8dB (f) RD-CTCF at SNR 8dB

FIGURE 9. **Confusion matrices for different techniques with one-layer LDA classifier.**

HTC, HOS and RD-CTCF perform outstandingly well at high and medium SNRs. However, in the range of $[-5, +10]$ dB, it can be seen that element HOS is slightly more sensitive to noise compared to other methods because the scalar features are easier to fluctuate.

3) INFLUENCE OF MODULATION TYPES

Different processing algorithms exhibit different noise immunity and differ in their ability to recognize different modulations. SCD, a classical representative of cyclostationary analysis, has a lower overall accuracy than other cyclic correlation spectra because it cannot discriminate high order QAMs and APSKs. However, it can distinguish between modulations such as ASK and FSK at low SNR because of the differences in peak size and amount [125]. As shown in Fig. 9, the other pre-processing confusion matrices at SNR -6dB are essentially a random chaotic distribution using an LDA classifier, while SCD and HOS can better discriminate ASK and BPSK signals. At SNR of 8dB, both kernel-based and high order cyclic correlation functions have high accuracy as in Fig. 9. Nevertheless, there are still some limitations on M-QAMs and M-APSKs tasks.

Element HOS is the simplest to compute without much accuracy loss, so many studies are based on this technique, still, the shortcoming is that it is composed of some numerical values, and some information may be lost in the calculation process. But cyclic temporal moment or cumulant functions (RD-CTCF) can compensate for this to some extent with better generalization capability. Moreover, it can be observed from Fig. 9 (b), (c) and (e) that high order algorithms can discriminate M-QAM better than other algorithms, but easily confuse within {32QAM, 128QAM} set and {16QAM, 64QAM, 256QAM} set, this is because they have the close distribution of statistical values as in Table 4. It can be seen that the latter algorithms in Fig. 8 generally perform more accurately than the former ones. On top of this, HTC and CCSD are more effective in discriminating amplitude modulations, while high order algorithms are more prominent for those involving phase modulation.

4) DISCUSSION

Our study shows that although the preprocessing methods studied differ in noise sensitivity and the ability to discriminate between modulation types, non-linear time-frequency transformations lead to better accuracy than linear ones. High order algorithms (HOS and RD-CTCF) can achieve good accuracy while maintaining relatively low computational complexity. The kernel-based cyclostationary techniques (HTC and CCSD) exhibit the best overall classification accuracy, but at higher computational cost and their ability to handle higher-order QAM and APSK can still be improved. Presented analysis shows that among the algorithms that perform more accurately, HOS is characterized by the lower order-of-magnitude required for inference time and, despite a slightly lower performance in the $[-5, +10]$ dB range,

is relatively competitive with more complex algorithms in lower and higher SNRs.

The state-of-the-art radio datasets [25] have contributed significantly to the field of AMC by providing a benchmark for evaluating deep learning approaches to AMC [126]. Although the classification accuracy for modulations such as OOK, M-ASK, BPSK, QPSK, GMSK, and OQPSK is excellent, improvements for higher-order M-APSK and M-QAM ($M \geqslant 64$), especially in low SNR scenarios remain a challenge. Most AMC systems are designed based on known modulation patterns, while in realistic scenarios it is possible that the modulation type was not present in the training set. There is little presented in the literature regarding this problem. Finally, real-time implementations in hardware are seldom published and this is still an open issue to be resolved for additional research.

VIII. CONCLUSION

This article highlights different signal representation techniques that can be applied to AMC. We have presented a number of common techniques and compared them under standardized conditions. In the presence of AWGN, some basic time-frequency transformations can effectively distinguish between FSK, PSK and other modulation types. HOS features are good in identifying M-QAM and M-PSK with low computational complexity. Second-order and generalizations to higher-order cyclostationary features perform well in diverse interference environments such as non-Gaussian noise and fading channels. Although the spectral correlation function did not perform well in our tests, the introduction of a Gaussian kernel leads to improved noise immunity over traditional linear time-frequency transforms, and a hyperbolic tangent kernel further improves robustness to impulse noise. While even simple techniques have good accuracy for low-order modulations or high SNR environments, the main challenge for AMC research lies in high-order modulations, low SNR environments, and coping with modulation types not known a priori.

In practice, the best approach depends on the particular application domain and environmental characteristics. However, in the general case of blind identification in unknown channel environments, the kernel function-based cyclostationary and high order cyclic temporal functions are likely to provide the best results. In addition, we provide a Python library that can reproduce the results in this paper and hope it can be used to improve future intelligent communication system design and software-defined radio applications.

REFERENCES

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

[2] J. L. Xu, W. Su, and M. Zhou, "Likelihood-ratio approaches to automatic modulation classification," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 4, pp. 455–469, Jul. 2011.

[3] C.-Y. Chi, C.-C. Feng, C.-H. Chen, and C.-Y. Chen, *Blind Equalization and System Identification: Batch Processing Algorithms, Performance and Applications*. New York, NY, USA: Springer, 2006.

[4] J. L. Xu, W. Su, and M. Zhou, "Software-defined radio equipped with rapid modulation recognition," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1659–1667, May 2010.

[5] L. Y. Uys, M. Gouws, J. J. Strydom, and A. S. J. Helberg, "The performance of feature-based classification of digital modulations under varying SNR and fading channel conditions," in *Proc. IEEE AFRICON*, Sep. 2017, pp. 198–203.

[6] O. A. Dobre, "Signal identification for emerging intelligent radios: Classical problems and new challenges," *IEEE Instrum. Meas. Mag.*, vol. 18, no. 2, pp. 11–18, Apr. 2015.

[7] D. H. Al-Nuaimi, I. A. Hashim, I. S. Z. Abidin, L. B. Salman, and N. A. M. Isa, "Performance of feature-based techniques for automatic digital modulation recognition and classification—A review," *Electronics*, vol. 8, no. 12, p. 1407, Nov. 2019.

[8] S. Peng, S. Sun, and Y.-D. Yao, "A survey of modulation classification using deep learning: Signal representation and data preprocessing," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 14, 2021, doi: 10.1109/TNNLS.2021.3085433.

[9] J. Walrand and P. P. Varaiya, *High-Performance Communication Networks*. San Mateo, CA, USA: Morgan Kaufmann, 2000.

[10] D. Wang, N. Zhang, Z. Li, F. Gao, and X. Shen, "Leveraging high order cumulants for spectrum sensing and power recognition in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1298–1310, Feb. 2018.

[11] W. Su, "Feature space analysis of modulation classification using very high-order statistics," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1688–1691, Sep. 2013.

[12] A. Abdelmutalab, K. Assaleh, and M. El-Tarhuni, "Automatic modulation classification based on high order cumulants and hierarchical polynomial classifiers," *Phys. Commun.*, vol. 21, pp. 10–18, Dec. 2016.

[13] O. A. Dobre, Y. Bar-Ness, and W. Su, "Higher-order cyclic cumulants for high order modulation classification," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2003, pp. 112–117.

[14] X. Zhou, Y. Wu, and B. Yang, "Signal classification method based on support vector machine and high-order cumulants," *Wireless Sensor Netw.*, vol. 2, pp. 48–52, Jan. 2010.

[15] L. Liu and J. Xu, "A novel modulation classification method based on high order cumulants," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Sep. 2006, pp. 1–5.

[16] G. A. Rovithakis, M. Maniadakis, and M. Zervakis, "A hybrid neural network/genetic algorithm approach to optimizing feature extraction for signal classification," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 34, no. 1, pp. 695–703, Feb. 2004.

[17] A. Smith, M. Evans, and J. Downey, "Modulation classification of satellite communication signals using cumulants and neural networks," in *Proc. Cognit. Commun. Aerosp. Appl. Workshop (CCAA)*, Jun. 2017, pp. 1–8.

[18] S. Majhi, R. Gupta, W. Xiang, and S. Glisic, "Hierarchical hypothesis and feature-based blind modulation classification for linearly modulated signals," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 11057–11069, Dec. 2017.

[19] G. Wang, H. Guo, G. Bi, S. K. Ting, and S. Razul, "Automatic modulation classification of cochannel communication signals based on cumulant," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[20] W. Chen, Y. Jiang, L. Zhang, and Y. Zhang, "A new modulation recognition method based on wavelet transform and high-order cumulants," *J. Phys., Conf.*, vol. 1738, no. 1, Jan. 2021, Art. no. 012025.

[21] O. A. Dobre, M. Oner, S. Rajan, and R. Inkol, "Cyclostationarity-based robust algorithms for QAM signal identification," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 12–15, Jan. 2012.

[22] P. Kruczek and J. Obuchowski, "Cyclic modulation spectrum—An online algorithm," in *Proc. 24th Medit. Conf. control Autom. (MED)*, Jun. 2016, pp. 361–365.

[23] J. Antoni, G. Xin, and N. Hamzaoui, "Fast computation of the spectral correlation," *Mech. Syst. Signal Process.*, vol. 92, pp. 248–277, Aug. 2017.

[24] N. Bidyanta, G. Vannhoy, M. Hirzallah, A. Akoglu, B. Ryu, and T. Bose, "GPU and FPGA based architecture design for real-time signal classification," in *Proc. Wireless Innov. Forum Conf. Wireless Commun. Technol. Softw. Defined Radio (WInnComm)*, Mar. 2015, pp. 70–79.

[25] (2018). *RF Datasets for Machine Learning*. DeepSig. [Online]. Available: https://www.deepsig.ai/datasets

[26] *The Python Language Reference*. Accessed: Sep. 20, 2022. [Online]. Available: https://docs.python.org/3/reference/index.html
[27] W. A. Gardner and C. M. Spooner, "The cumulant theory of cyclostationary time-series. I. Foundation," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3387–3408, Dec. 1994.
[28] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. 1st Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.
[29] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 2015.
[30] J. Ma, S.-C. Lin, H. Gao, and T. Qiu, "Automatic modulation classification under non-gaussian noise: A deep residual learning approach," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[31] S. Huang, C. Lin, W. Xu, Y. Gao, Z. Feng, and F. Zhu, "Identification of active attacks in Internet of Things: Joint model- and data-driven automatic modulation classification approach," *IEEE Internet Things J.*, vol. 8, no. 3, pp. 2051–2065, Feb. 2021.
[32] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2595–2621, 4th Quart., 2018.
[33] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
[34] L. Mingzhu, Z. Yue, S. Lin, and D. Jingwei, "Research on recognition algorithm of digital modulation by higher order cumulants," in *Proc. 4th Int. Conf. Instrum. Meas., Comput., Commun. Control*, Sep. 2014, pp. 686–690.
[35] Y. Zhao, Y.-T. Xu, H. Jiang, Y.-J. Luo, and Z.-W. Wang, "Recognition of digital modulation signals based on high-order cumulants," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2015, pp. 1–5.
[36] D. Das, P. K. Bora, and R. Bhattacharjee, "Cumulant based automatic modulation classification of QPSK, OQPSK, 8-PSK and 16-PSK," in *Proc. 8th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Jan. 2016, pp. 1–5.
[37] A. Ali and F. Yangyu, "Higher-order statistics based modulation classification using hierarchical approach," in *Proc. IEEE Adv. Inf. Manage., Communicates, Electron. Autom. Control Conf. (IMCEC)*, Oct. 2016, pp. 370–374.
[38] U. Satija, M. S. Manikandan, and B. Ramkumar, "Performance study of cyclostationary based digital modulation classification schemes," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–5.
[39] M. H. Valipour, M. M. Homayounpour, and M. A. Mehralian, "Automatic digital modulation recognition in presence of noise using SVM and PSO," in *Proc. 6th Int. Symp. Telecommun. (IST)*, Nov. 2012, pp. 378–382.
[40] W. Zhang, "Automatic modulation classification based on statistical features and support vector machine," in *Proc. 31th URSI Gen. Assem. Sci. Symp. (URSI GASS)*, Aug. 2014, pp. 1–4.
[41] S. A. Ghunaim, Q. Nasir, and M. A. Talib, "Deep learning techniques for automatic modulation classification: A systematic literature review," in *Proc. 14th Int. Conf. Innov. Inf. Technol. (IIT)*, Nov. 2020, pp. 108–113.
[42] B. Kim, J. Kim, H. Chae, D. Yoon, and J. W. Choi, "Deep neural network-based automatic modulation classification technique," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2016, pp. 579–582.
[43] H. Wang and L. Guo, "A new method of automatic modulation recognition based on dimension reduction," in *Proc. Forum Cooperat. Positioning Service (CPGPS*, May 2017, pp. 316–320.
[44] A. Hazza, M. Shoaib, S. A. Alshebeili, and A. Fahad, "An overview of feature-based methods for digital modulation classification," in *Proc. 1st Int. Conf. Commun., Signal Process., Appl. (ICCSPA)*, Feb. 2013, pp. 1–6.
[45] F. Wang, S. Huang, H. Wang, and C. Yang, "Automatic modulation classification exploiting hybrid machine learning network," *Math. Problems Eng.*, vol. 2018, pp. 1–14, Dec. 2018.
[46] W. Jiang, X. Wu, Y. Wang, B. Chen, W. Feng, and Y. Jin, "Time–frequency-analysis-based blind modulation classification for multiple-antenna systems," *Sensors*, vol. 21, no. 1, p. 231, Jan. 2021.
[47] R. M. Hidalgo, J. G. Fernandez, R. R. Rivera, and H. A. Larrondo, "A simple adjustable window algorithm to improve FFT measurements," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 1, pp. 31–36, Feb. 2002.
[48] O. S. Mossad, M. ElNainay, and M. Torki, "Deep convolutional neural network with multi-task learning scheme for modulations recognition," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 1644–1649.
[49] P. K. H. L and L. Shrinivasan, "Automatic digital modulation recognition using minimum feature extraction," in *Proc. 2nd Int. Conf. Comput. Sustain. Global Develop. (INDIACom)*, Mar. 2015, pp. 772–775.
[50] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018, doi: 10.1109/ACCESS.2018.2818794.
[51] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 961–1005, Sep. 1990.
[52] M. S. Ford, "The illustrated wavelet transform handbook: Introductory theory and applications in science," *Health Phys.*, vol. 84, no. 5, pp. 667–668, May 2003.
[53] V. Pukhova, E. Gorelova, G. Ferrini, and S. Burnasheva, "Time-frequency representation of signals by wavelet transform," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, 2017, pp. 715–718.
[54] Y. Lee, K. Hirakawa, and T. Q. Nguyen, "Lossless compression of CFA sampled image using decorrelated Mallat wavelet packet decomposition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2721–2725.
[55] S. A. Saleh, M. A. Khan, and M. A. Rahman, "Application of a wavelet-based MRA for diagnosing disturbances in a three-phase induction motor," in *Proc. 5th IEEE Int. Symp. Diag. Electr. Mach., Power Electron. Drives*, Sep. 2005, pp. 1–6.
[56] P. S. Addison, "Wavelet transforms and the ECG: A review," *Physiol. Meas.*, vol. 26, no. 5, pp. R155–R199, Oct. 2005.
[57] K. C. Ho, W. Prokopiw, and Y. T. Chan, "Modulation identification of digital signals by the wavelet transform," *IEE Proc.-Radar, Sonar Navigat.*, vol. 147, no. 4, pp. 169–176, Aug. 2000.
[58] L. Hong and K. C. Ho, "Identification of digital modulation types using the wavelet transform," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, vol. 1, Nov. 1999, pp. 427–431.
[59] X. Wei and Z. Cao, "Fast identification of amplitude modulated signals at low SNR," in *Proc. IEEE Int. Symp. Microw., Antenna, Propag. EMC Technol. Wireless Commun.*, vol. 2, Aug. 2005, pp. 1119–1122.
[60] L.-L. Meng and X.-J. Si, "An improved algorithm of modulation classification for digital communication signals based on wavelet transform," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, vol. 3, Nov. 2007, pp. 1226–1231.
[61] A. Vyas and J. Paik, "Review of the application of wavelet theory to image processing," *IEIE Trans. Smart Process. Comput.*, vol. 5, no. 6, pp. 403–417, Dec. 2016.
[62] M. B. Abdulkareem, "Design and development of multimodal medical image fusion using discrete wavelet transform," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Apr. 2018, pp. 1629–1633.
[63] P. Sun and J. Qin, "Analysis of impulse noise based on wavelet transform for military applications," in *Proc. IEEE AUTOTEST*, Sep. 2014, pp. 183–187.
[64] S. Hassanpour, A. M. Pezeshk, and F. Behnia, "A robust algorithm based on wavelet transform for recognition of binary digital modulations," in *Proc. 38th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2015, pp. 508–512.
[65] J. Chen, Y. Kuo, J. Li, F. Fu, and Y. Ma, "Digital modulation identification by wavelet analysis," in *Proc. 6th Int. Conf. Comput. Intell. Multimedia Appl. (ICCIMA)*, Aug. 2005, pp. 29–34.
[66] L. Wang, S. Guo, and C. Jia, "Recognition of digital modulation signals based on wavelet amplitude difference," in *Proc. 7th IEEE Int. Conf. Softw. Eng. Service Sci. (ICSESS)*, Aug. 2016, pp. 627–630.
[67] R. G. Stockwell, L. Mansinha, and R. P. Lowe, "Localization of the complex spectrum: The S transform," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 998–1001, Apr. 1996.
[68] L. Mansinha, R. G. Stockwell, and R. P. Lowe, "Pattern analysis with two-dimensional spectral localisation: Applications of two-dimensional S transforms," *Phys. A Statist. Mech. Appl.*, vol. 239, nos. 1–3, pp. 286–295, May 1997.
[69] Y. Lin, X. Xu, B. Li, J. Pang, and R. Zhou, "The feature extraction and classification for signals based on the s-transform," in *Proc. Int. Conf. Meas., Inf. Control*, vol. 1, 2013, pp. 550–553.
[70] U. Satija, M. Mohanty, and B. Ramkumar, "Automatic modulation classification using S-transform based features," in *Proc. 2nd Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Feb. 2015, pp. 708–712.
[71] Z. Zhao, S. Wang, W. Zhang, and Y. Xie, "A novel automatic modulation classification method based on stockwell-transform and energy entropy for underwater acoustic signals," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Aug. 2016, pp. 1–6.

[72] E. Azzouz and A. K. Nandi, *Automatic Modulation Recognition of Communication Signals*. New York, NY, USA: Springer, 2013.
[73] A. Swami and B. M. Sadler, "Hierarchical digital modulation classification using cumulants," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 416–429, Mar. 2000.
[74] P. Marchand, C. Le Martret, and J.-L. Lacoume, "Classification of linear modulations by a combination of different orders cyclic cumulants," in *Proc. IEEE Signal Process. Workshop Higher-Order Statist.*, Jul. 1997, pp. 47–51.
[75] C. M. Spooner, "On the utility of sixth-order cyclic cumulants for RF signal classification," in *Proc. 35th Asilomar Conf. Signals, Syst. Comput.*, vol. 1, Nov. 2001, pp. 890–897.
[76] G. Zhou, G. B. Giannakis, and A. Swami, "On polynomial phase signals with time-varying amplitudes," *IEEE Trans. Signal Process.*, vol. 44, no. 4, pp. 848–861, Apr. 1996.
[77] D. Hatzinakos, "Blind deconvolution channel identification and equalization," in *Digital Signal Processing Systems: Implementation Techniques* (Control and Dynamic Systems), vol. 68. New York, NY, USA: Academic, 1995, pp. 279–331.
[78] P. R. Halmos, *Naive Set Theory*. New York, NY, USA: Dover, 2017.
[79] J. Renard, J. Verlant-Chenet, J.-M. Dricot, P. De Doncker, and F. Horlin, "Higher-order cyclostationarity detection for spectrum sensing," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, no. 1, pp. 1–10, Dec. 2010.
[80] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, "Cyclostationarity-based modulation classification of linear digital modulations in flat fading channels," *Wireless Pers. Commun.*, vol. 54, no. 4, pp. 699–717, Sep. 2010.
[81] A. Swami, G. B. Giannakis, and G. Zhou, "Bibliography on higher-order statistics," *Signal Process.*, vol. 60, no. 1, pp. 65–126, Jul. 1997.
[82] S. Hakimi and A. Ebrahimzadeh, "Digital modulation classification using the bees algorithm and probabilistic neural network based on higher order statistics," *Int. J. Inf. Commun. Technol. Res.*, vol. 7, no. 2, pp. 1–15, Dec. 2015.
[83] Z. Zhu, M. W. Aslam, and A. K. Nandi, "Support vector machine assisted genetic programming for MQAM classification," in *Proc. Int. Symp. Signals, Circuits Syst. (ISSCS)*, Jun. 2011, pp. 1–6.
[84] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Process.*, vol. 86, no. 4, pp. 639–697, 2006.
[85] C. Spooner. *Symmetries of Second-Order Probabilistic Parameters in CSP*. Accessed: Sep. 20, 2021. [Online]. Available: https://cyclostationary.blog/2019/12/12/symmetries-of-second-order-probabilistic-parameters-in-csp/
[86] E. L. D. Costa, "Detection and identification of cyclostationary signals," Ph.D. dissertation, Dept. Elect. Eng., Naval Postgraduate School, Monterey, CA, USA, 1996.
[87] R. S. Roberts, W. A. Brown, and H. H. Loomis, "Computationally efficient algorithms for cyclic spectral analysis," *IEEE Signal Process. Mag.*, vol. 8, no. 2, pp. 38–49, Apr. 1991.
[88] W. Gardner, "Measurement of spectral correlation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1111–1123, Oct. 1986.
[89] P. Borghesani and J. Antoni, "A faster algorithm for the calculation of the fast spectral correlation," *Mech. Syst. Signal Process.*, vol. 111, pp. 113–118, Oct. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327018301869
[90] D. Vucic, "Cyclic spectral analysis of OFDM/QAM modulation using stochastic matrix-based method," *Facta Univ. Electron. Energetics*, vol. 16, no. 3, pp. 343–353, 2003.
[91] H. Zhang, D. Le Ruyet, and M. Terré, "Spectral correlation of multicarrier modulated signals and its application for signal detection," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, pp. 1–14, Dec. 2009.
[92] G. J. Mendis, J. Wei, and A. Madanayake, "Deep learning-based automated modulation classification for cognitive radio," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Dec. 2016, pp. 1–6.
[93] X. Zhu, Y. Lin, and Z. Dou, "Automatic recognition of communication signal modulation based on neural network," in *Proc. IEEE Int. Conf. Electron. Inf. Commun. Technol. (ICEICT)*, Aug. 2016, pp. 223–226.
[94] L. Zhao, H. Xu, and C. Bai, "Blind identification of the number of sub-carriers for orthogonal frequency division multiplexing-based elastic optical networking," *Opt. Commun.*, vol. 411, pp. 101–107, Mar. 2018.
[95] A. I. R. Fontes, A. de M. Martins, L. F. Q. Silveira, and J. C. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Syst. Appl.*, vol. 42, pp. 1–8, Jan. 2015.
[96] A. V. Dandawate and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2355–2369, Sep. 1994.
[97] A. I. R. Fontes, J. B. A. Rego, A. M. de Martins, L. F. Q. Silveira, and J. C. Principe, "Cyclostationary correntropy: Definition and applications," *Expert Syst. Appl.*, vol. 69, pp. 110–117, Mar. 2017.
[98] J. C. Principe and D. Xu, "An introduction to information theoretic learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 3, Jul. 1999, pp. 1783–1787.
[99] T. Liu, T. Qiu, and S. Luan, "Hyperbolic-tangent-function-based cyclic correlation: Definition and theory," *Signal Process.*, vol. 164, pp. 206–216, Nov. 2019.
[100] T. V. R. O. Camara, A. D. L. Lima, B. M. M. Lima, A. I. R. Fontes, A. D. M. Martins, and L. F. Q. Silveira, "Automatic modulation classification architectures based on cyclostationary features in impulsive environments," *IEEE Access*, vol. 7, pp. 138512–138527, 2019.
[101] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives*. New York, NY, USA: Springer, 2010.
[102] J. Ma and T. Qiu, "Automatic modulation classification using cyclic correntropy spectrum in impulsive noise," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 440–443, Apr. 2019.
[103] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[104] S. M. Hiremath, S. Behura, S. Kedia, S. Deshmukh, and S. K. Patra, "Deep learning-based modulation classification using time and stockwell domain channeling," in *Proc. Nat. Conf. Commun. (NCC)*, Feb. 2019, pp. 1–6.
[105] H. Wu, Y. Li, L. Zhou, and J. Meng, "Convolutional neural network and multi-feature fusion for automatic modulation classification," *Electron. Lett.*, vol. 55, no. 16, pp. 895–897, Aug. 2019.
[106] F. Wang, C. Yang, S. Huang, and H. Wang, "Automatic modulation classification based on joint feature map and convolutional neural network," *IET Radar, Sonar Navigat.*, vol. 13, no. 6, pp. 998–1003, Jun. 2019.
[107] Z. Zhang, C. Wang, C. Gan, S. Sun, and M. Wang, "Automatic modulation classification using convolutional neural network with features fusion of SPWVD and BJD," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 5, no. 3, pp. 469–478, Sep. 2019.
[108] D. Wang, M. Zhang, Z. Li, J. Li, M. Fu, Y. Cui, and X. Chen, "Modulation format recognition and OSNR estimation using CNN-based deep learning," *IEEE Photon. Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, Oct. 1, 2017.
[109] J. Lee, B. Kim, J. Kim, D. Yoon, and J. W. Choi, "Deep neural network-based blind modulation classification for fading channels," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2017, pp. 551–554.
[110] A. Ali and F. Yangyu, "Automatic modulation classification using deep learning based on sparse autoencoders with nonnegativity constraints," *IEEE Signal Process. Lett.*, vol. 24, no. 11, pp. 1626–1630, Nov. 2017.
[111] W. Xie, S. Hu, C. Yu, P. Zhu, X. Peng, and J. Ouyang, "Deep learning in digital modulation recognition using high order cumulants," *IEEE Access*, vol. 7, pp. 63760–63766, 2019.
[112] S. H. Lee, K.-Y. Kim, J. H. Kim, and Y. Shin, "Effective feature-based automatic modulation classification method using dnn algorithm," in *Proc. Int. Conf. Artif. Intell. Inf. Commun. (ICAIIC)*, 2019, pp. 557–559.
[113] W. Shi, D. Liu, X. Cheng, Y. Li, and Y. Zhao, "Particle swarm optimization-based deep neural network for digital modulation recognition," *IEEE Access*, vol. 7, pp. 104591–104600, 2019.
[114] A. Ozen and C. Ozturk, "A novel modulation recognition technique based on artificial bee colony algorithm in the presence of multipath fading channels," in *Proc. 36th Int. Conf. Telecommun. Signal Process. (TSP)*, Jul. 2013, pp. 239–243.
[115] M. W. Aslam, Z. Zhu, and A. K. Nandi, "Automatic modulation classification using combination of genetic programming and KNN," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2742–2750, Jun. 2012.
[116] Y. Wang, M. Liu, J. Yang, and G. Gui, "Data-driven deep learning for automatic modulation recognition in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074–4077, Apr. 2019.
[117] M. Zhang, Y. Zeng, Z. Han, and Y. Gong, "Automatic modulation recognition using deep learning architectures," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.

[118] N. Wang, Y. Liu, L. Ma, Y. Yang, and H. Wang, "Multidimensional CNN-LSTM network for automatic modulation classification," *Electronics*, vol. 10, no. 14, p. 1649, Jul. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/14/1649
[119] P. Goncalves and R. G. Baraniuk, "Pseudo affine Wigner distributions: Definition and kernel formulation," *IEEE Trans. Signal Process.*, vol. 46, no. 6, pp. 1505–1516, Jun. 1998.
[120] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Hum. Eugenics*, vol. 7, no. 2, pp. 179–188, Aug. 2012.
[121] C. Duan, Y. Zhan, and H. Liang, "An identification technique for the co-frequency mixed communication signals based on cumulants," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–9, Dec. 2018.
[122] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, and S. J. Van Der Walt, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," *Nature Methods*, vol. 17, no. 3, pp. 261–272, 2020.
[123] G. R. Lee, R. Gommers, F. Wasilewski, K. Wohlfahrt, and A. O'Leary, "PyWavelets: A Python package for wavelet analysis," *J. Open Source Softw.*, vol. 4, no. 36, p. 1237, 2019.
[124] T. Oliphant. (2006). *NumPy: A Guide to NumPy*. USA: Trelgol Publishing. [Online]. Available: http://www.numpy.org/
[125] W. Gardner, W. Brown, and C.-K. Chen, "Spectral correlation of modulated signals: Part II-digital modulation," *IEEE Trans. Commun.*, vol. COM-35, no. 6, pp. 595–601, Jun. 1987.
[126] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

XUEYUAN LIU received the B.E. degree in electrical and information engineering from The University of Sydney, Australia, in 2019, where she is currently pursuing the Ph.D. degree with full scholarship. Her research interests include artificial intelligence, radio frequency systems, and FPGA design.

CAROL JINGYI LI (Student Member, IEEE) received the M.E. degree in electrical engineering from The University of Melbourne, in 2016. She is currently pursuing the Ph.D. degree with The University of Sydney, Australia, with full scholarship. Her current research interests include wireless communication and reconfigurable computing.

CRAIG T. JIN received the Ph.D. degree in electrical engineering from The University of Sydney, in 2001. He is an Associate Professor and the Head of the Computing and Audio Research Laboratory, The University of Sydney; and also the Founder of VAST Audio Pty. Ltd. He is the author or coauthor of more than 50 journals or conference papers and holds six patents. His research interests include spatial audio and neuromorphic engineering. He has received national recognition in Australia for his invention of a spatial hearing aid.

PHILIP H. W. LEONG (Senior Member, IEEE) received the B.Sc., B.E., and Ph.D. degrees from The University of Sydney. In 1993, he was a Consultant with ST Microelectronics, Italy. From 1997 to 2009, he was with The Chinese University of Hong Kong. He is currently a Professor of computer systems with the School of Electrical and Information Engineering, The University of Sydney; a Visiting Professor with Imperial College London; a Chief Technology Advisor with ClusterTech; and a Chief Technology Officer with CruxML. He was the Co-Founder and the Program Co-Chair of the International Conference on Field Programmable Technology (FPT) and the Program Co-Chair of the International Conference on Field Programmable Logic and Applications (FPL), and is the Senior Associate Editor of the *ACM Transactions on Reconfigurable Technology and Systems*.