

Received June 4, 2020, accepted July 31, 2020, date of publication August 27, 2020, date of current version September 8, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3019243

Dyadic Aggregated Autoregressive Model (DASAR) for Automatic Modulation Classification

MARCO ANTONIO PINTO-ORELLANA¹ AND HUGO LEWI HAMMER^{2,3}

¹Department of Mechanical, Electronics and Chemical Engineering, Faculty of Technology, Art and Design, Oslo Metropolitan University, 0103 Oslo, Norway

²Department of Information Technology, Faculty of Technology, Art and Design, Oslo Metropolitan University, 0103 Oslo, Norway

³Simula Metropolitan Center, Oslo Metropolitan University, 0103 Oslo, Norway

Corresponding author: Marco Antonio Pinto-Orellana (marcop@oslomet.no)

This work was supported by the Research Council of Norway to Project 273599, ‘Patient-Centric Engineering in Rehabilitation (PACER).’

ABSTRACT In this article, we presented a novel spectral estimation method, the dyadic aggregated autoregressive model (DASAR), that characterizes the spectrum dynamics of a modulated signal. DASAR enhances automatic modulation classification (AMC) on environments where new or unknown modulation techniques are introduced, and only size-restricted data is accessible to train classification algorithms. A key component for obtaining efficient machine learning-based classification is the development of valuable knowledge-descriptive features. DASAR constructs a multi-level spectral representation by subdividing a signal into successive dyadic segments where each partition is modeled as an aggregation of single-frequency autoregressive processes. Thus, the model ensures a robust representation at the segment level, while the multi-level decomposition can capture time-varying spectra. As a feature extraction model, DASAR can provide useful learning features related to signals with complex spectra. The effectiveness of our model was tested on a dataset comprised of 11 different modulation techniques and realistic transmission medium characteristics. Using only 200 128-point samples per modulation scheme (1% of the available signal samples) and a proper selection of a classification algorithm, DASAR reaches accuracy up to 70.96% compared with a maximum accuracy of 43.62% using the state-of-art methods tested under the same conditions.

INDEX TERMS Automatic modulation classification, signal processing, spectrum modeling, autoregressive model, machine learning.

I. INTRODUCTION

A reliable communication system enforces that a transmitted message will only be understood if the transmitter and receiver have complete information about the communication parameters, including the type and characteristics of the desired modulation (i.e., the techniques to adapt the message to the transmission medium conditions). However, it is feasible to infer some properties about the communication channel’s configuration if we intercept the modulated message passing through the channel. This inference is the central goal in automatic modulation classification (AMC) [1]. AMC was developed for military applications such as electronic

warfare, surveillance, and threat analysis [2]. Nevertheless, its relevance made it fundamental for other widespread civilian purposes, including software-defined radio (SDR) and spectrum management. SDR is the base of self-organized communication infrastructures that rely on AMC to recognize the modulation in real-time and to adapt or optimize resources during the ongoing transmission process. The rapid progress and spread of high-speed wireless communication networks (especially 5G and WiFi 802.11ax), and the consequent spectrum saturation, impulsive the role of AMC in spectrum management allowing the regulatory entities to detect misused spectrum ranges by non-authorized communication channels.

AMC algorithms can be classified into likelihood-based (LB-AMC) and feature-based (FB-AMC) methods [1]–[4].

The associate editor coordinating the review of this manuscript and approving it for publication was Michael Lyu.

LB-AMC estimates the likelihood of a finite set of modulation schemes given the observed signals. Multiple likelihood ratios are commonly applied as the main discrimination method [3]. It is granted that LB-AMC methods can theoretically achieve optimal solutions (if prior information of the transmitter is available). However, due to the complexity of those likelihood functions, the calculations depend on Markov chain Monte Carlo [4] techniques that are intrinsically time computationally expensive and inhibit the application of LB-AMC for embedded systems.

Moreover, FB-AMC techniques search for suboptimal alternatives and provide reasonable and flexible solutions with less computational requirements that do not depend on prior knowledge. FB-AMC was conventionally based on the development of expert features based on the nature of the modulation process: 5th, and 8th-order cyclic cumulants [4], Fourier transform of the received signal [5] or the phase signal [6]. However, the detection accuracy of those methods is limited. In recent research, there was an interest in the development of advanced machine learning methods to increase that accuracy. In this trend, several models were proposed based on deep learning: deep belief networks [7], extreme learning machines [8], stacked autoencoders [1], or convolutional neural networks [9]–[13].

The state-of-art methods mentioned above have reached high levels of accuracy in classifying the modulation type. In some cases, accuracies higher than 90% were reported [8], [12], [13]. However, there are still some challenges, mainly the need for larger datasets for training with all desired and possible modulation techniques in order to train the neural networks properly.

In this article, we propose a preventive solution for those cases when a new modulation scheme, or channel condition, is introduced, and there are only a limited modulated signal samples. For this objective, we proposed to extract spectral features under a classical machine learning classification model. The features are generated from a novel stochastic time-spectral decomposition of the signal magnitude and phase. Our decomposition model, the dyadic aggregated autoregressive model (DASAR), is based on a dyadic (binary) partition of the signals, where the spectrum of each individual segment is modeled by a set of autoregressive models. DASAR can offer noise-robust mechanisms of spectral autoregressive representations while allowing to represent complex dynamic and varying spectra. Therefore, our model can also capture the spectral signal dynamics while retaining only the most significant spectral components (likely to contain the modulated and carrier signal). Furthermore, in order to ensure the computational feasibility of this method, we also incorporated a dynamic estimation method.

The rest of this article is organized as follows: Section 2 describes the generalized signal model, the aggregated autoregressive model, the dyadic decomposition technique, and the estimation algorithm; section 3 illustrates the dataset and the experimental conditions; in section 4, the results of the classification algorithms are shown along

with the respective discussions. Finally, we present our conclusions in section 5.

II. DASAR MODEL

In a common communication system, a carrier signal, with amplitude A and carrier frequency f_c : $c(t) = A \cos(2\pi f_c t)$ is modified according to a source signal $x(t)$ in order to produce a signal $s(t)$ prepared to the conditions of the communication channel (baseband signal). In essence, three major parameters are manipulated from the carrier signal: frequency, phase, and amplitude. We can summarize the diverse single-carrier modulation techniques through the accompanying equation:

$$s(t) = A (g(t))^{\kappa_{PAM}} (x(t))^{\kappa_{AM}} \cdot \cos\left(2\pi t \left(f_c + \kappa_{FM} \Delta f \int_0^t x(\tau) d\tau\right) + \kappa_{PM} x(t)\right) \quad (1)$$

where κ_{AM} , κ_{FM} , κ_{PM} , κ_{PAM} are modulation scheme parameters and $g(t)$ determines the shape of the pulse (in digital communications), typically a cosine frequency shaping filter:

$$g(t) = \text{sinc}\left(\frac{t}{T}\right) \frac{\cos\left(\pi \alpha \frac{t}{T}\right)}{1 - \left(2\alpha \frac{t}{T}\right)^2} \quad (2)$$

controlled by the roll-off factor $\alpha \in [0, 1]$.

Now, let us consider the parameter vector:

$$\kappa_\theta = (\kappa_{AM}, \kappa_{FM}, \kappa_{PM}, \kappa_{PAM}) \quad (3)$$

Combinations of this vector generate the majority of existing single-carrier modulation techniques characterized by [2]. For instance, amplitude modulation double-sided band (AM DSB) and amplitude shift keying (ASK) are the analog and digital case when $\kappa_\theta = (1, 0, 0, 0)$. Frequency modulation (FM) and frequency shift keying (FSK) are a particular case of Equation 1 with $\kappa_\theta = (0, 1, 0, 0)$. Phase modulation (PM) and phase shift keying (PSK), phase schemes in analog and digital communications, are special cases of Equation 1 when $\kappa_\theta = (0, 0, 1, 0)$. Furthermore pulse amplitude modulation (PAM) is obtained when $\kappa_\theta = (0, 0, 0, 1)$; and quadrature amplitude modulation (QAM) under $\kappa_\theta = (1, 0, 1, 0)$.

We should recognize that the source signal $s(t)$ will suffer alterations during its transmission over a wired or wireless medium, and the signal received will be different. Therefore, we can model the measured signal $r(t)$ at the receiver as:

$$r(t) = \omega(t) + \alpha e^{j(2\pi f_0 t + \theta_0)} \cdot \int_{-\infty}^{\infty} s(\tau) p(\tau - t) h(t - \tau + \epsilon_T) d\tau \quad (4)$$

where α is the channel gain, $p(\cdot)$ is the pulse shape of the channel, $h(\cdot)$ is the channel response, and ϵ_T is the symbol timing error, and f_0 is the carrier frequency that can be delayed by θ_0 radians. Moreover, it is also often observed that the channel will be affected by a thermal noise interference $\omega(t)$ (modeled with an additive Gaussian noise [14]) inherently related to the maximum signal-to-noise ratio supported by the

channels (SNR). We refer to [2] and [15] for further details about this model.

If the pulse shape and channel response are known, the receiver is able to mitigate the distortion impact of the channel as it takes place in communication with sufficient knowledge between transmitter and receiver. In those cases, the model can be simplified as [2]:

$$r(t) = \alpha e^{j(2\pi f_0 t + \theta_0)} s(t) + \omega(t) \quad (5)$$

A. SPECTRAL REPRESENTATION USING AGGREGATED AUTOREGRESSIVE COMPONENTS

For complex modulation schemes, the encoded and transmitted symbols (Equation 1 and 5) can be recovered applying discrete inverse Fourier transform [16]. It is implicit in this process that time-located frequency variations for each modulated symbol can be ignored. Nevertheless, Fourier transform is still suitable to demodulate the signals under some constraints. In this article, we proposed a parametric method to model the signal spectrum; therefore, the complete spectrum can be represented by a reduced set of parameters while spurious components are rejected.

The crucial element in DASAR is the spectral representation as an aggregated autoregressive model (AAR). For efficient parameter estimation and simple representation, in this study we use aggregated second-order autoregressive model AAR(2). However, this choice does not restrict the application of DASAR to higher-order AAR models. An autoregressive model of second order AR(2) describes a real and finite signal $x(t)$ (sampled at a fixed interval $1/f_s$) as the stochastic process generated by a linear combination of the two previous points, $x(t-1)$ and $x(t-2)$, and an additive Gaussian error term $\varepsilon(t)$:

$$x(t) = \phi_1 x(t-1) + \phi_2 x(t-2) + \varepsilon(t) \quad (6)$$

where ϕ_1 and ϕ_2 are known as the autoregressive coefficients. In this case, without loss of generality, we assume that $\mathbb{E}[X(t)] = 0$. The error component $\varepsilon(t) \sim \mathcal{WN}(0, \sigma_\varepsilon^2)$, is a zero-mean random variable, with variance σ_ε^2 , and serially uncorrelated $\mathbb{E}[\varepsilon(t)\varepsilon(t-\tau)] = 0 \forall \tau$.

For stationary AR(2) processes, the power spectral density (PSD) $S_x(\omega)$ is characterized by:

$$S_x(\omega) = \frac{\sigma_\varepsilon^2}{|1 - \phi_1 e^{-2j\pi\omega} - \phi_2 e^{-4j\pi\omega}|^2} \quad (7)$$

where ω is the normalized frequency with respect to the sampling frequency f_s : $\omega = \frac{f}{f_s}$.

We should note that $S_x(\omega)$ has a single maximum at the dominating frequency ω^* :

$$\omega^* = \frac{1}{2\pi} \arccos\left(\frac{\phi_1(\phi_2 - 1)}{4\phi_2}\right) = \frac{1}{2\pi} \arctan\left(\frac{\phi_2}{\phi_1}\right) \quad (8)$$

when $\phi_1^2 + 4\phi_2 < 0$. Given our interest in the spectral information contained in an AR process, we can reformulate the autoregressive coefficients based on ω^* and a

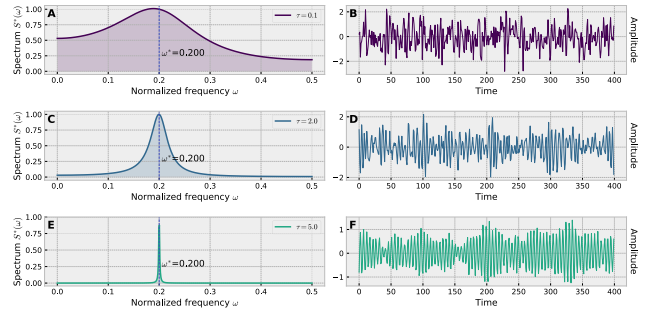


FIGURE 1. Second-order autoregressive processes as stochastic harmonic oscillators controlled by τ and ω^* .

parameter τ :

$$\phi_1 = \frac{2}{1 + e^{-\tau}} \cos(2\pi\omega^*) \quad (9)$$

$$\phi_2 = -\frac{1}{(1 + e^{-\tau})^2} \quad (10)$$

AR(2) processes can effectively model stochastic oscillations with a frequency around ω^* [17]. In these models, τ controls the randomness of the central oscillation frequency: small values can model signals with frequency components widely spread around ω^* (Figure 1.A-B); and large values of τ can model signals with an clear central frequency (Figure 1.D-E).

However, the received signal $r(t)$ can have more than one main resonating frequency (this is more noticeable in AM/ASK), and therefore a single AR(2) would be insufficient to represent the spectrum of the signal accurately. Nevertheless, we can take advantage of the concise stochastic frequency representation offered by AR(2), establishing an aggregated model as the superposition of several AR(2) components with distinct parameters. Some theoretical properties of these types of models were introduced by Chong *et al.* [18] and generalized by Dacunha-Castelle *et al.* for AR(p) processes [19].

An aggregated model AAR(p, K) is defined as the sum of K uncorrelated components where each one is characterized through an AR(p) process:

$$y(t) = \sum_{k=1}^K z_k(t) + v(t) \quad (11)$$

where ε_k is the additive white noise $\varepsilon_k \sim \mathcal{WN}(0, \sigma_{\varepsilon_k}^2)$, and each $z_k(t)$ is a latent unobserved AR(p) time series, in our target case, an AR(2) process associated with a central frequency ω_k^* , and a frequency randomness τ_k (Equation 9, 10 and 7):

$$z_k(t) = \frac{\cos(2\pi\omega_k^*)}{1 + e^{-\tau_k}} x(t) - \frac{1}{(1 + e^{-\tau_k})^2} z_k(t-1) + \varepsilon_k(t) \quad (12)$$

Given that $z_i(t)$ and $z_j(t)$ are uncorrelated for $i \neq j$, the spectrum of an AAR (2, K) is determined by

$$S_y(\omega) = \sum_{k=1}^K \frac{\sigma_{\varepsilon k}^2}{\left| 1 - \frac{\cos(2\pi\omega_k^*)}{1+e^{-\tau_k}} e^{-2j\pi\omega} + \frac{1}{(1+e^{-\tau_k})^2} e^{-4j\pi\omega} \right|^2} + \sigma_v^2 \quad (13)$$

B. ESTIMATION METHOD

The inference of the set of parameters $\{\phi_{1,1}, \phi_{2,1}, \dots, \phi_{1,K}, \phi_{2,K}, \sigma_1^2, \dots, \sigma_K^2\}$ of an AAR (2, p) model is still an open problem. Wong *et al.* proposed Kalman filters with expectation-maximization as a method for estimating AAR (2, 4) [20]. Gao *et al.* also proposed to Kalman filter with a least squares estimation based on block resampling for AAR (2, 4) [21]. It should be emphasized that in the latter, it is required apriori knowledge of the target signal, and a reasonable fixed resonating frequencies $\{\omega_k^*\}$. To the best of our knowledge, there is no general method that provides an estimation method without relying on strong assumptions related to the number of representative components of the signal.

We introduce an estimation method that automatically infers a pertinent number of components (and their parameters) that can explain a set of observations. Given a signal that has a fixed, and unknown, number of dominating frequencies with at least a separation of Δf , we proposed an estimation algorithm composed of three stages:

- 1) *Estimation of an interpolated fast Fourier transform (FFT) of a signal $x(t)$.* The autocovariance function of a zero-mean stationary process, $x[t]$, is given by $\gamma_X(\ell) = \mathbb{E}[x[t]x[t-\ell]]$, where $\mathbb{E}[\cdot]$ is the expectation operator and the spectrum is further given by the Fourier transform:

$$S_X(\omega) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} e^{-j\ell\omega} \gamma_X(\ell) \quad (14)$$

For a realization (or observation) of this process, the FFT squared magnitude of the recorded signal can be used as an estimate of $S_X(\omega)$. Moreover, to increase the frequency scale of the digitalized FFT signal, we can interpolate the estimates in the frequency domain with a Dirichlet kernel; or equivalently, zero-padding the signal in the time domain,

$$x_Z = \begin{cases} x(t) & t \leq T \\ 0 & t > T. \end{cases}$$

Thus, the initially estimated spectrum is

$$S^{(0)}(\omega) = S_{FFT}(\omega) = \frac{1}{T} \left| \sum_{t=1}^{T+B} x_Z(t) \exp(-j\pi t\omega) \right| \quad (15)$$

- 2) *Fit an AR(2) model that describes the dominating frequency in ω^* .* Let w be the frequency where $S^{(k)}$ has a maximum value (consider that initially, $k = 0$). Then,

restrict the neighborhood of w with a Gaussian window $B(\omega; \omega^*)$ around it with a standard deviation equal to the frequency separation Δf :

$$B(\omega; \omega^*) = \exp\left(-\left(\frac{\omega - \omega^*}{\Delta f}\right)^2\right) \quad (16)$$

Now, given the L_2 distance function defined by

$$L_2(\omega; \omega^*, \tau) = \sum_{\omega} \left(S^{(k)}(\omega) B(\omega; \omega^*) - S(\omega; \omega^*, \tau) \right)^2 \quad (17)$$

define the estimate $\hat{\tau}$ as the nonlinear least square of B , i.e., the minimum value of $L(\omega)$:

$$\tau^* = \arg \min_{\tau} (L(\omega; \omega^*, \tau)) \quad (18)$$

- 3) *Calculate the residuals.* Obtain the unexplained spectrum for a next iteration $k + 1$:

$$S^{(k+1)}(\omega) = \left[S^{(k)}(\omega) - S(\omega; \omega^*, \tau^*) \right]^+ \quad (19)$$

- 4) The previous steps should be repeated until convergence ($\sum_{\omega} |S^{(k+1)}(\omega)| < \varepsilon$), or until reaching the maximum number of desired components.

C. DYADIC TIME PARTITION

AAR (2, k) offers a model that explains the most significant frequency components in the signal, but ignoring their time origin. This absent time-location information is highly relevant in dynamic modulated signals because of the modulated signal's dynamic nature. Therefore, inspired by the decomposition in the discrete wavelet transform, the dyadic aggregated autoregressive model (DASAR) addressed this time-location issue through a dyadic (or binary) decomposition of the time series where a signal segment on n -th level is split into two segments ($n + 1$ -th level). For each segment, we fit a AAR (2, k) model with a maximum of k components.

This decomposition structure provides a representation of the frequency components that are allowed to vary according to time during the complete recorded transmission. Thus, stable components such as the carrier frequencies are likely to be captured in the first level (level 0 in Figure 2) due to their stability and magnitude compared with the modulated components and the noise. Subsequent levels can capture components that are most significant only in specific intervals, or can capture minor variations in the carrier frequencies as they are common for frequency modulation.

Figure 2 shows an example of a decomposition of signals with 11 different modulation schemes into 4 levels. Note that compared with the decomposition applied in wavelet transform, DASAR also keeps the high-frequency dominating components at each level.

III. DATASET

In order to ensure the repeatability of the results presented in this study, we utilized a publicly available dataset for assessing the performance of our model: the RadioML2016.10a

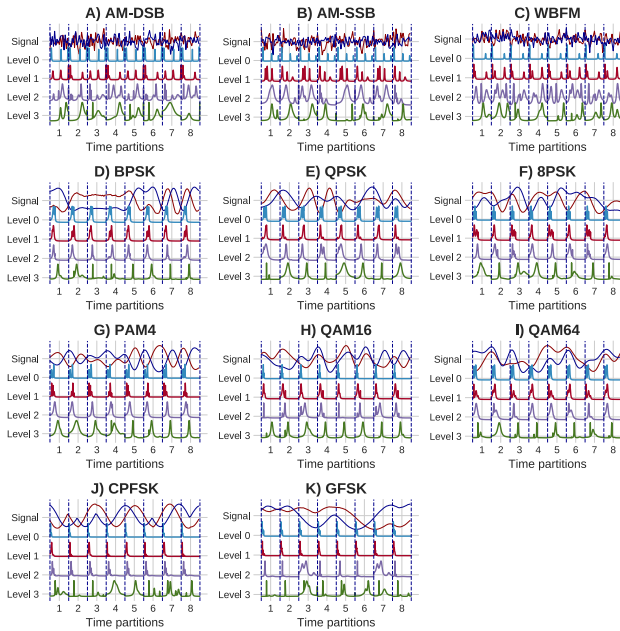


FIGURE 2. 4-level DASAR decomposition of the different 11 modulation schemes presented in the RadioML2016.10a dataset.

(RML) dataset. This is a synthetic database created using the software GNU Radio to simulate several realistic transmission conditions. It was originally introduced by O’Shea [14]. In RML, the communication channel has an average sampling rate of 200KHz with a standard deviation of 0.01Hz. The effect of the channel on the carrier (carrier frequency drift) is simulated through a zero-mean Gaussian random variable with a standard deviation of 0.01Hz. Moreover, Doppler effects (common in wireless transmissions) are simulated with stochastic frequency shifts up to 1Hz. Finally, channel delays are simulated through a fractional delay model, with three fractional tap delays: 0, 0.8 and 1.7 taps.

RML comprises 220000 in-phase/quadrature samples with a length of 128 points each one, those segments simulate several noisy conditions with signal-to-noise ratios (SNRs) ranging from -20dB to 18dB . Note that the overall RML simulation parameters (SNR levels, maximum Doppler shift, and channel tap delays) can simulate mobile communications in indoor and outdoor environments with communication devices at varying speeds and under different interference levels. Nevertheless, no burst errors are included in the simulated communication channels. Moreover, RML comprises 11 modulation schemes including 3 analog methods: double-sideband amplitude modulation (AM-DSB), single-sideband amplitude modulation (AM-SSB), wideband frequency modulation (WBFM); as well as eight digital modulation techniques: 2-, 4- and 8-level phase shift keying (BPSK, QPSK, 8PSK), pulse-amplitude modulation (PAM), 16- and 64-level quadrature amplitude modulation (QAM16, QAM64), Gaussian and continuous-phase frequency-shift keying (GFSK, CPFSK). For a detailed explanation about the simulated modulation techniques and their respective parameters, we refer to [14] and [9].

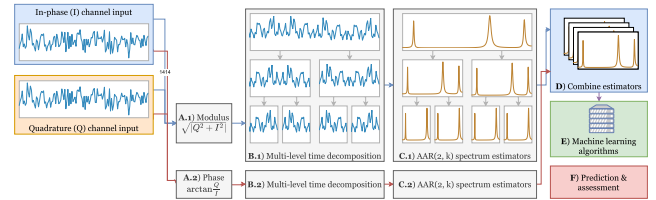


FIGURE 3. Feature extraction method for AMC based on the DASAR model.

A. METHOD

The fundamental intention of this study is to introduce a design for those instances where we have minimal access to samples of an unusual set of channel conditions or unforeseen modulation settings. These circumstances can occur when proprietary industrial multi-carrier modulation (MCM) protocols are introduced, or abnormal time-varying electromagnetic propagation patterns are found, such as Lim *et al.* measured in the coastal area of South Korea [22]. For the cases where a large dataset is available, O’Shea *et al.*, Rejendram *et al.*, and Ramjee *et al.* have reported accuracies closed to 90% using convolutional neural networks (CNN) [9], long short-term memory neural networks (LSTM) [13], and convolutional long short-term deep neural networks (CLDNN) [12], respectively.

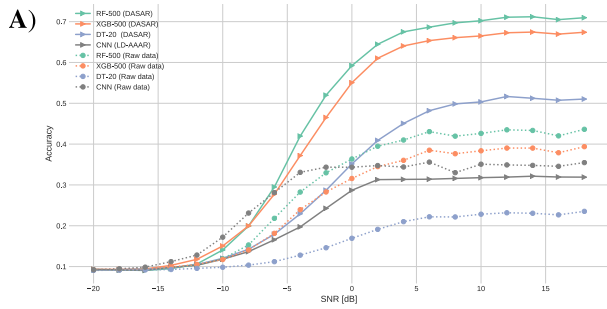
We extracted features using a 4-level DASAR model with a maximum of 4 components per segment (Figure 3) such that, at least, 8 data points are processed in the lowest levels given the size of the sample segments. The spectral information is extracted from the modulus and phase of the samples in RML, following the parameter estimation procedure described in Section II.

We adopt a training set composed of a randomly chosen 1% of the dataset (2200 samples) with an SNR higher than 0, and the rest of the dataset as testing set (217800 samples). Due to computational constraints, this process was only repeated 30 times. We should emphasize that this proportion of 1%-99% allowed us to replicate a condition of limited knowledge of the available types of modulations using less than 200 sample segments per modulation scheme.

We also contrast two conditions: our processed features and raw data using four machine learning classifiers: the CNN architecture introduced by [14] as a reference algorithm, random forest of 500 trees (RF500), extreme gradient boosting trees with 500 trees (XGB500), a decision tree with 20 maximum levels (DT20).

IV. RESULTS AND DISCUSSION

It is usually that the AMC discriminators displayed the classifier’s accuracy as a function of the signal-to-noise ratio of the modulated signal [9], [12], [13], in order to observe their ability to detect the correct modulation method in environments with high, medium and low interference. It is also frequent that the displayed accuracy is normalized with respect to the number of samples in each SNR level [9], [12], [13].



B)

SNR	DASAR			
	RF-500	XGB-500	DT-20	CNN
-20	9.220	9.320	9.190	9.320
-18	9.170	9.310	9.240	9.170
-16	9.040	9.410	9.410	9.100
-14	9.490	10.290	9.680	9.810
-12	10.630	11.770	10.470	10.310
-10	14.060	15.050	12.030	11.800
-8	19.930	19.960	14.240	13.660
-6	29.490	27.730	17.920	16.570
-4	41.960	37.190	22.920	19.720
-2	51.980	46.510	28.700	24.250
0	59.220	55.050	35.150	28.690
2	64.450	61.030	40.900	31.300
4	67.500	64.050	45.060	31.350
6	68.650	65.350	48.190	31.400
8	69.700	66.070	49.840	31.590
10	70.220	66.500	50.340	31.780
12	71.080	67.280	51.660	31.920
14	71.200	67.440	51.230	32.140
16	70.490	66.950	50.740	31.940
18	70.960	67.400	51.040	31.910

SNR	Raw data			
	RF-500	XGB-500	DT-20	CNN
-20	9.200	9.230	9.060	9.210
-18	9.230	9.250	9.120	9.460
-16	9.370	9.370	9.170	9.850
-14	9.650	9.880	9.290	11.210
-12	10.360	10.520	9.560	12.820
-10	11.910	11.730	9.830	17.200
-8	15.290	14.070	10.350	23.110
-6	21.810	18.140	11.200	28.120
-4	28.250	23.940	12.800	33.080
-2	32.970	28.270	14.610	34.350
0	36.360	31.580	16.920	34.340
2	39.480	34.490	19.120	34.750
4	41.000	36.030	21.000	34.390
6	43.060	38.500	22.190	35.590
8	41.950	37.640	22.150	33.000
10	42.620	38.360	22.830	35.080
12	43.500	39.030	23.180	34.900
14	43.350	39.020	23.060	34.790
16	42.020	37.880	22.670	34.560
18	43.620	39.370	23.520	35.470

FIGURE 4. Normalized classifiers' accuracies with respect to SNRs: A) curve, and B) detailed table.

The accuracy-per-SNR curve of the eight simulated configurations (raw data and DASAR with four classification algorithms) is displayed in Figure 5. Initially, we can observe that the performance of the features extracted from DASAR using RF500 showed the best performance with a detection accuracy higher than 70% with SNRs higher than 10dB.

We acknowledge that O'Shea's CNN model [9] will provide low performance (lower than DT20 in our simulations),

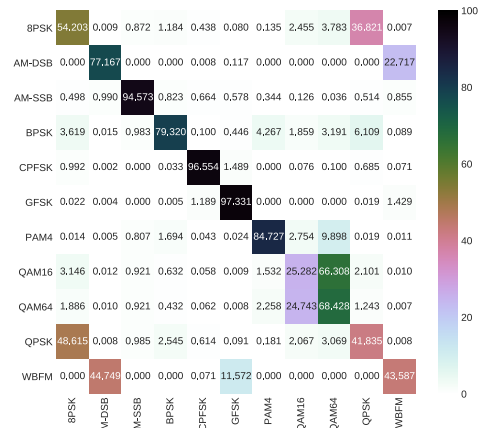


FIGURE 5. Normalized confusion matrix of RF500 trained with DASAR features. The matrix is constructed over the performance of the classifier on signals with SNR higher than zero. Note that the values are shown in percentages.

due to the limitation of the training data size. As it was also denoted by [13], a deep learning architecture will require more than 50% of the dataset input for reaching an accuracy level of 90%.

We also present the normalized confusion matrix obtained by the RF500 classifier in Figure 5. This structure allowed us to identify the central reasons that illustrate the performance of the method:

- The modulation schemes that involve one or more distinct central frequencies were accurately recognized. Thus, AM-SSB has a prediction accuracy of 94.6%, while CPFSK and GFSK have accuracies higher than 96.6%. The latter techniques modulate discrete input signals shifting the carrier frequency, and the RF500 classifier accurately identifies this change.
- However, gradual variations in the carrier frequency were not detected by the classification algorithm (WBFM was correctly classified only in 43.6% of the cases). We hypothesized that DASAR with a maximum of 4 components per level was not sufficient to represent this continuous change in frequency accurately, and the variations were improperly interpreted as wide fluctuations over the carrier wave, and therefore, labeled as amplitude modulation AM-DSB (that produces the same precise behavior in frequency). Note that 22.7% of the AM-DSB segments are recognized as WBFM, and 44.5% of the WBFM samples are classified as AM-DSB.
- Phase transformations cannot be fully described by the PSD, and therefore, by a limited-level DASAR. We should note that this type of modulation has higher bit error rates, given its high sensitivity to changes in the medium [2]. In our simulations, we observe a confusion of the RF500 classifier between 8-PSK and QPSK, and QAM16 with respect to QAM64. The

first set (QPSK and 8-PSK) corresponds to the same modulation technique with 4 and 8 levels, respectively. In both, the digital input information is encoded by only changing the phase of the signal. The second pair (QAM16 and QAM64) are also two variants of a quadrature amplitude modulation that encodes digital signals through combinations of amplitude and phases. Even though the limitations of an AAR(2) spectral representation, the dyadic structure of DASAR can represent the transitions caused by the variations of phase and allow RF500 to identify the algorithms. But, a minor difference in the type of phase transitions (e.g., from QAM16 to QAM64) limits the effectiveness of a 4-level DASAR for differentiating the correct scheme.

V. CONCLUSION

We presented a spectrum-estimation model to serve as a feature-based method for automatic modulation classification with an emphasis in environments with limited access to samples of the modulation schemes. Our method, the dyadic aggregated autoregressive model (DASAR), starts performing a dyadic decomposition, i.e., it divides a signal based on successive binary partitions. Each segment is modeled as a sum of second-order autoregressive processes from which a robust estimator of the spectrum can be obtained. The inference of the parameters in the later processes is performed using an efficient and flexible numerical method.

As a feature extraction method for AMC, DASAR was tested on the RadioML2016.10a dataset, which contains a large sample set of modulation techniques under realistic transmission conditions. As the state-of-art proposed for AMC, we explored a convolutional neural network (CNN) architecture with raw data as a classification method, as long as decision trees, and high-performant ensemble trees (random forests and extreme gradient boosting trees) trained with DASAR features. Later, we used a train/test split technique to evaluate the performance of those techniques, repeated 30 times. To assess the generalization power of our method, we simulate a restrictive environment condition where only a minimal portion of the complete set is accessible, in such a way that 2200 samples (1% of the dataset) with SNR higher than 0dB were randomly selected for training the algorithms, and 217800 samples (99% of the dataset) is utilized for testing the performance.

Random forest classifiers with DASAR-based features showed accuracies higher than 70% for identifying 11 different modulation techniques when the signal-to-noise ratio was higher than 10dB. In comparison, the state-of-art methods showed an accuracy lower than 35.5%. Surprisingly, DASAR showed an accuracy higher than the state-of-art (59%) even when the noise and signal have the same power level (0dB). We should emphasize that alternative state-of-art deep-learning methods have reported higher accuracies when larger datasets are available. Nevertheless, the observed

results revealed promising further applications of DASAR for environments where novel modulation schemes are introduced, or extensive datasets cannot be provided or constructed.

REFERENCES

- [1] A. Ali, F. Yangyu, and S. Liu, "Automatic modulation classification of digital modulation signals with stacked autoencoders," *Digit. Signal Process.*, vol. 71, pp. 108–116, Dec. 2017.
- [2] Z. Zhu and A. K. Nandi, *Automatic Modulation Classification: Principles, Algorithms and Applications*. Hoboken, NJ, USA: Wiley, 2015.
- [3] N. E. Lay and A. Polydoros, "Per-survivor processing for channel acquisition, data detection and modulation classification," in *Proc. 28th Asilomar Conf. Signals, Syst. Comput.*, Oct. 1994, pp. 1169–1173.
- [4] 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, p. 137, 2007.
- [5] Z. Yu, Y. Q. Shi, and W. Su, "M-ary frequency shift keying signal classification based-on discrete Fourier transform," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2003, pp. 1167–1172.
- [6] P. C. Sapiiano, J. D. Martin, and R. J. Holbeche, "Classification of PSK signals using the DFT of phase histogram," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, May 1995, pp. 1868–1871.
- [7] 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.
- [8] A. Güner, F. Alçin, and A. Şengür, "Automatic digital modulation classification using extreme learning machine with local binary pattern histogram features," *Measurement*, vol. 145, pp. 214–225, Oct. 2019.
- [9] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Engineering Applications of Neural Networks*. Cham, Switzerland: Springer, 2016, pp. 213–226.
- [10] S. Zhou, Z. Yin, Z. Wu, Y. Chen, N. Zhao, and Z. Yang, "A robust modulation classification method using convolutional neural networks," *EURASIP J. Adv. Signal Process.*, vol. 2019, no. 1, Mar. 2019.
- [11] 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.
- [12] S. Ramjee, S. Ju, D. Yang, X. Liu, A. E. Gamal, and Y. C. Eldar, "Fast deep learning for automatic modulation classification," 2019, *arXiv:1901.05850*. [Online]. Available: <https://arxiv.org/abs/1901.05850>
- [13] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin, "Deep learning models for wireless signal classification with distributed low-cost spectrum sensors," *IEEE Trans. Cognit. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.
- [14] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Convolutional radio modulation recognition networks," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, 2016, pp. 213–226.
- [15] L. Rugini and P. Banelli, "BER of OFDM systems impaired by carrier frequency offset in multipath fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2279–2288, Sep. 2005.
- [16] H. Minn, V. K. Bhargava, and K. B. Letaief, "A robust timing and frequency synchronization for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 24, no. 5, pp. 822–839, May 2003.
- [17] R. B. Govindan, S. Kota, T. Al-Shargabi, A. N. Massaro, T. Chang, and A. du Plessis, "Effect of electrocardiogram interference on cortico-cortical connectivity analysis and a possible solution," *J. Neurosci. Methods*, vol. 270, pp. 76–84, Sep. 2016.
- [18] T. T.-L. Chong and K.-T. Wong, "Time series properties of aggregated AR(2) processes," *Econ. Lett.*, vol. 73, no. 3, pp. 325–332, Dec. 2001.
- [19] D. Dacunha-Castelle and L. J. Fermin, "Aggregation of autoregressive processes and long memory," 2008, *arXiv:0811.1917*. [Online]. Available: <https://arxiv.org/abs/0811.1917>
- [20] K. F. K. Wong, A. Galka, O. Yamashita, and T. Ozaki, "Modelling non-stationary variance in EEG time series by state space GARCH model," *Comput. Biol. Med.*, vol. 36, no. 12, pp. 1327–1335, Dec. 2006.

- [21] X. Gao, W. Shen, B. Shahbaba, N. J. Fortin, and H. Ombao, "Evolutionary state-space model and its application to time-frequency analysis of local field potentials," *Statistica Sinica*, vol. 30, pp. 1561–1582, Jul. 2020.
- [22] T. H. Lim, S. Wang, Y. Chong, Y. B. Park, J. Ko, and H. Choo, "High altitude ducts causing abnormal wave propagation in coastal area of Korea," *Microw. Opt. Technol. Lett.*, vol. 62, no. 2, pp. 643–650, Oct. 2019.



MARCO ANTONIO PINTO-ORELLANA received the M.Sc. degree in computer science from the Federal University of Viçosa, Brazil, in 2017, and the Diploma degree in satellite design and space segment from the China Academy of Space Technology, Beijing, China. He is currently pursuing the Ph.D. degree in engineering with Oslo Metropolitan University (OsloMet), Norway. His research interests include signal processing, game theory, and biomedical time series.



HUGO LEWI HAMMER received the M.Sc. and Ph.D. degrees from the Norwegian University of Science and Technology, Trondheim, Norway, in 2003 and 2008, respectively. He holds a position of a Professor in statistics with the Department of Computer Science, Oslo Metropolitan University (OsloMet), Norway. His current research interests include computer intensive statistical methods, Bayesian statistics, and learning systems.

...