Toro Betancur, Veronica; Carmona Valencia, Augusto; Marulanda Bernal, José Ignacio

Signal Detection and Modulation Classification for Satellite Communications

Published in:
Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning

DOI:
10.1145/3432291.3432297

Published: 22/10/2020

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Toro Betancur, V., Carmona Valencia, A., & Marulanda Bernal, J. I. (2020). Signal Detection and Modulation Classification for Satellite Communications. In *Proceedings of the 2020 3rd International Conference on Signal Processing and Machine Learning* (pp. 114–118). Article 3432297 ACM.
https://doi.org/10.1145/3432291.3432297

Signal detection and modulation classification for satellite communications

Verónica Toro-Betancur
veronica.torobetancur@aalto.fi
Aalto University, Finland

Augusto Carmona Valencia
acarmon2@eafit.edu.co
Universidad EAFIT, Colombia

José Ignacio Marulanda Bernal
jmarulan@eafit.edu.co
Universidad EAFIT, Colombia

ABSTRACT

Amateur ground stations are gaining increasing importance as both academic and hobby activities. However, due to the limited energy resources available in amateur satellites, ground stations need to be located in isolated places in order to establish a reliable communication. This usually implies limited Internet access. Hence, ground stations need to be able to recognize incoming signal without completely relying on an Internet connection. For this reason, we propose an algorithm to estimate parameters such as amplitude, center frequency, bandwidth and modulation type for amateur radio applications. For signal detection, we use an absolute-valued sinc approximation which estimates the center frequency and bandwidth of signals with signal-to-noise ratios over -6 dB with a precision of 5% and 2% respectively. In addition, Support Vector Machines (SVM) binary classifiers are used in series to classify the four most common modulation types used in amateur satellites. With accuracies over 90%, SVM outperforms solutions based on Artificial Neural Networks.

CCS CONCEPTS

• **Computing methodologies** → **Supervised learning**; • **Applied computing** → *Aerospace.*

KEYWORDS

Modulation Classification, Signal Detection, Spectrum Sensing, Support Vector Machines, Artificial Neural Networks, Amateur Satellites

ACM Reference Format:

Verónica Toro-Betancur, Augusto Carmona Valencia, and José Ignacio Marulanda Bernal. 2020. Signal detection and modulation classification for satellite communications. In *Proceedings of SPML'20*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnn.nnnnn

1 INTRODUCTION

The current number of active satellites orbiting the Earth is around 2218, from which approximately 150 are amateur satellites [2, 7, 8]. This type of satellites use unlicensed frequency bands for uplink and downlink transmissions. Such bands include VHF (Very high frequency), UHF (Ultra high frequency) and, in some few cases, the S (2.4 GHz) band [2]. By using unlicensed bands, amateur satellites allow amateur radio operators to decode their transmissions using their own ground stations. However, to be able to decode such transmissions, some information has to be available for the ground stations, e.g., carrier frequency and modulation type. This information is usually publicly available in multiple websites [7, 11]. Nevertheless, some ground stations are located in difficult access areas where there is no Internet connectivity. In such cases, a ground station with offline information about satellites will be quickly unable to track them and decode their signals. Even if the antenna is able to capture a satellite signal, it will not know with certainty its source, i.e., the satellite that generated it. As a result, the ground station does not have the required information to decode the signal [13].

To address this problem, we propose an algorithm to detect signals from unknown satellites and determine their amplitude, center frequency, bandwidth and modulation type in order to proceed with its demodulation. Particularly, our algorithm uses an *absolute-valued sinc* approach to model signal features such as amplitude, center frequency and bandwidth with no previous information of it. While for the modulation classification, a comprehensive study on feature extraction is presented and evaluated under two Machine Learning (ML) techniques. Namely, Artificial Neural Networks (ANN) and Support Vector Machines (SVM). We consider the four most common modulation types in amateur radio: BPSK (Binary Phase-Shift Keying), BFSK (Binary Frequency-shift Keying), CW (Continuous Waveform) with a Morse code encoding and GMSK (Gaussian minimum-shift keying). Such modulations are present in 13.4%, 54.6%, 30.6% and 23.3% of the active amateur satellites, respectively. Note that the sum of the given percentages is higher than 100% since satellites usually support multiple modulations.

The contributions of this paper are as follows: *(i)* we propose an algorithm for signal detection and modulation classification targeting, specifically, amateur satellite signals. *(ii)* we present a comprehensive study on feature extraction for the modulation classification problem. Such an study is, to the best of our knowledge, the first one for amateur satellite communications. *(iii)* we test our proposed algorithm in both simulated and real data.

This paper is organized as follows. Section 2 presents the related work in signal detection and modulation classification. Section 3 details the models used to approximate the signal features, while Section 4 provides the methods used for modulation classification. Section 5 shows the performance evaluation of the algorithm when applied to simulated and real data. Finally, Section 6 gives some concluding remarks and future work.

SPML'20, October 22–24, 2020, Beijing, China,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnn.nnnnn

2 RELATED WORK

Signal detection techniques have been widely used in cognitive radios to detect transmissions from primary users in a specific frequency band. Namely, energy detection [1], matched filtering [6], and cyclostationarity detection [12] are the most common techniques. Such methods, however, do not provide information about the signal features, instead, they only determine the presence of signals in a certain frequency band.

On the other hand, Automatic Modulation Classification (AMC) has also been extensively addressed in the literature [5]. For instance, in [3] ANN are used to classify seven different modulation types. While [10] exploits deep learning techniques to classify both digital and analog modulations. [4] uses the constellation diagram of signals to classify different levels of QAM (Quadrature Amplitude Modulation), ASK (Amplitude-Shift Keying), and PSK modulations. Although all these approaches present AMC solutions, none of them target satellite communications, which typically have low SNR (signal-to-noise ratio) values due to the low-power nature of small satellites such as nano and picosatellites. In contrast, our proposed algorithm is able to successfully classify this type of signals while also providing signal detection and parameter estimation.

The work in [14] address ACM and signal detection in satellite communications. However, such work focuses on PSK, APSK and QAM modulations. Instead, our work also targets the CW and FSK-type modulations which are widely used in amateur satellites.

3 SIGNAL DETECTION

The parameters of the received signal are estimated by approximating its amplitude, center frequency and bandwidth. Such approximation is done by modeling the FFT (Fast Fourier Transform) of the signal with an absolute-valued *sinc* function, as proposed in [15]. The parameters of the *sinc* function are adjusted in an iterative fashion until the best fit is found. The models used for each parameter are discussed in the following.

3.1 Amplitude model

The amplitude of the *sinc* function is adjusted by minimizing the mean squared error between the signal FFT (V_l) and the *sinc* approximation (S_l). However, only the data contained in the main lobe of the *sinc* are used. Eq. 1 shows the objective function for the amplitude approximation.

$$\min(E) = \min\left(\sum_{l\epsilon L} \left[V_l - \Delta_A S_l\left(x_l\right)\right]^2\right), \tag{1}$$

where L is a set that contains all of the elements in the main lobe of the *sinc* and Δ_A is the amplitude of the *sinc* function. Note that Δ_A is the parameter that must be optimized. Deriving Eq. 1 respect to Δ_A and optimizing, we obtain Eq. 2.

$$\Delta_A = \frac{\sum_{l\epsilon L}\left[V_l S_l\left(x_l\right)\right]}{\sum_{l\epsilon L}\left[S_l\left(x_l\right)\right]^2}. \tag{2}$$

The amplitude of the *sinc* is updated iteratively following Eq. 2. This process runs until an error restriction is reached.

3.2 Frequency model

The initial frequency approximation (x_{off}) is done at the point where the maximum value of the FFT is located, as in Eq. 3.

$$x_{\text{off}} = \text{argmax}(\text{FFT}). \tag{3}$$

For the frequency calculation, we use the model presented in [15] which consist of a torques balance. In this model, the areas under the right-side (M_R) and the left-side (M_L) of the *sinc* are to be balanced. Such areas are calculated as in Eq. 4.

$$M_{L,R} = \frac{1}{L}\sum_{l\epsilon L_{L,R}}\left[V_l - S\left(x_l\right)\right], \tag{4}$$

where

$$L_R = \{l : l \epsilon L, l > x_{\text{off}}\},\ L_L = \{l : l \epsilon L, l \leq x_{\text{off}}\}. \tag{5}$$

Then, M_R and M_L are compared with each other and the offset is adjusted accordingly to balance the areas. The offset is moved only one unit at a time to avoid divergence in the subroutine. This subroutine runs until the offset returns to the same point it was in the previous iteration.

3.3 Bandwidth model

For the bandwidth calculation, the main lobe of the *sinc* is approximated with the function shown in Eq. 6, as proposed in [15].

$$Q\left(\alpha, x\right) = A - A\left(\frac{\Delta_\alpha}{L}\right)^2\left(x - x_{\text{off}}\right)^2, \tag{6}$$

where $Q\left(\alpha, \varphi\right)$ is defined as in [15].

Then, the error for the values contained in the main lobe are defined as in Eq. 7

$$E = \sum_{l\epsilon L}\left(x_l - \varphi_l\right)^2. \tag{7}$$

With the previous error, Δ_α can be calculated for the bandwidth update, as follows

$$\Delta_\alpha = \frac{\sum_{l\epsilon L}\left[\Gamma_l^2\right]}{\sum_{l\epsilon L}\left[\left(x_l - x_{offset}\right)\mu\Gamma_l^4\right]}, \tag{8}$$

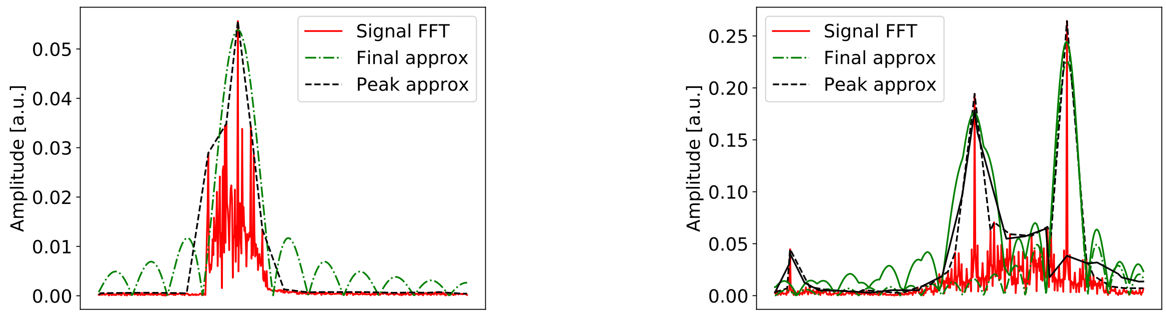
where

$$\Gamma_l = L\sqrt{\frac{A - V_l}{A}}. \tag{9}$$

This subroutine runs 50 times and a record of the mean squared error is stored in order to determine the value of Δ_α that minimizes the error.

Each of these models for amplitude, frequency and bandwidth run as a subroutine in the algorithm published as open source code[1]. In such algorithm, the incoming data is first transformed by FFT, then a peak approximation is done, which is performed by linear interpolations between two peaks in a window (this approximation is shown in Figure 1). Later, the amplitude, frequency and bandwidth subroutines run until the error conditions are reached. Then, if the error condition for the whole signal FFT window is not accomplished, a new approximation is done and the algorithm returns

[1] https://github.com/VeronicaToro/Signal-identifier



(a) BPSK signal **(b) BFSK signal**

Figure 1: Signal FFT with a peak and a *sinc* approximation

a double-approximation, which is represented by two different *sinc* functions (Figure 1b). The double-approximation approach is indeed necessary for BFSK signals, as they contain two main peaks in its FFT. Finally, the algorithm returns the computed parameters.

4 MODULATION CLASSIFICATION

Four modulation types were considered for the modulation classification problem: BPSK, BFSK, CW and GMSK. To address such a problem, we consider Support Vector Machines (SVM) with a linear kernel and Artificial Neural Networks (ANN) with different parameters to finally choose the method and the set of parameters that achieves the highest accuracy. On the other hand, to effectively choose the set of features that allow accurate classifications in any of the ML methods, we used the *Scikit-learn* univariate feature selection algorithm[2] to visualize the features that highly affect the classification.

We evaluate the following 8 features: (i) the standard deviation (as shown in Eq. 10) of both, (ii) the instantaneous frequency (Eq. 11) and (iii) the instantaneous phase (Eq. 12) of the signal. The correlation coefficient (ρ) between the signal and either (iv) a BFSK signal (ρ_f), (v) a BPSK signal (ρ_p), (vi) a sine wave of the same frequency as the signal (ρ_s), or (vii) a one-period sine wave (ρ_s'). And (viii) the mean amplitude of the signal ($\overline{A}$). Note that the reference signals used in features (iv) - (vii) are noise-free simulated signals.

$$\sigma = \sqrt{\frac{1}{L}\left[\sum \phi^2(i)\right] - \left[\sum \phi(i)\right]^2}, \tag{10}$$

$$\phi_f(t) = \frac{1}{2\pi}\frac{d\left[\arg\left(z(t)\right)\right]}{dt}, \tag{11}$$

$$\phi_p(t) = \text{unwrap}\left[\arg\left(z(t)\right)\right], \tag{12}$$

where

$$\arg\left(z(t)\right) = \arctan\left(\frac{\text{Im}\{z\}}{\text{Re}\{z\}}\right). \tag{13}$$

In the previous equations, the Hilbert transform of the signal $z(t)$ was used, which has a real (Re$\{z\}$) and an imaginary part (Im$\{z\}$). Table 1 shows the obtained scores for each of the evaluated features in this *base case*.

[2] https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.html

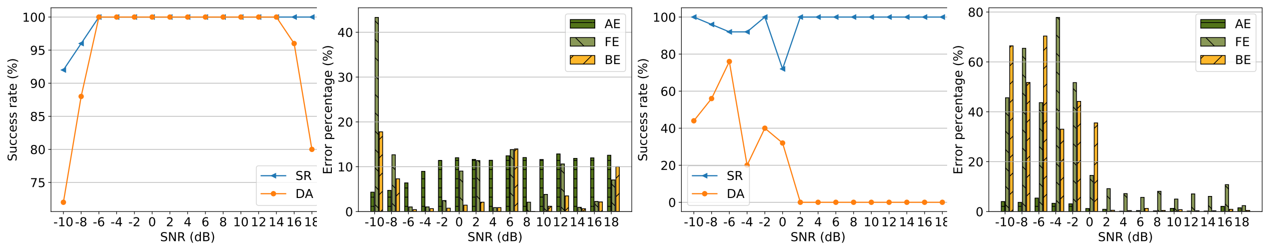
Feature number	Feature	Score
3	ϕ_p	17254.75
4	ρ_f	237.72
6	ρ_s	66.42
5	ρ_p	12.44
8	$\overline{A}$	9.89
7	ρ_s'	7.98
2	ϕ_f	1.51
1	σ	0.12

Table 1: Scores of the evaluated features.

The selected features have different importance for each modulation. For instance, the difference between the mean amplitude $\overline{A}$ of a CW signal is highly different to that of any other modulation. However, this difference is not important when you compare a BFSK and and GMSK signal. For this reason, we also evaluated the scores of each feature when only two modulation types are considered. As a result, we obtained the scores for each pair of modulations. We call this classification problem, the *binary case*. Table 2 shows the two features with the highest scores for each set of modulations.

Modulations	Feature number	Feature	Score
BPSK, BFSK	3	ϕ_p	15867.17
	4	ρ_f	89.35
BPSK, CW	3	ϕ_p	7971.31
	5	ρ_p	6.67
BPSK, GMSK	3	ϕ_p	2660.47
	6	ρ_s'	2.93
BFSK, CW	3	ϕ_p	1421.78
	5	ρ_f	78.87
BFSK, GMSK	3	ϕ_p	5723.50
	8	ρ_f	94.19
CW, GMSK	3	ϕ_p	1458.53
	7	$\overline{A}$	4.48

Table 2: Features with highest scores in the binary case.



(a) BFSK (b) BFSK (c) BPSK (d) BPSK

Figure 2: Success rate and error percentage for BFSK and BPSK signals.

To determine the modulation type of a test signal in the binary case, such signal is passed through each binary classifier. At the end, the correct modulation type is the one with the highest aggregated probability. This is similar to the voting process in a random forest approach.

In Section 5.2, we evaluate the accuracy of both: the base case and the binary case.

5 PERFORMANCE EVALUATION

The performance of the signal detection algorithm is evaluated in terms of error percentage, i.e., the difference between the approximations made and the real values, divided by the real value. On the other hand, the modulation classification methods are assessed by their accuracy. In order to evaluate such metrics, we used and modified the algorithms in [9] to simulate signals with different SNRs, varying from -10 dB to 18 dB. For each SNR value, 25 datasets were created with 1024 samples each, at a sample rate of 32 kHz. The following sections shows the results for the used approaches.

5.1 Signal detection

The performance of the signal detection algorithm for BFSK and BPSK signals can be seen in Figure 2. Specifically, the success rate (SR) and percentage of double-approximation (DA) are shown in Figures 2a and 2c, respectively. For the tested datasets, the SR refers to the number of times the algorithm converged to a solution, while DA indicates the number of times the algorithm performed a second approximation. Clearly, a double-approximation is expected from a BFSK signal, while it is undesirable for a BPSK signal. In fact, it can be seen that for the tested BFSK signal, double-approximations are made in 100% of the cases already at -6 dB. However, this percentage drops at high SNR due to the low content of noise, i.e., after the first approximation is made, the root mean square error is lower than the threshold and thus, no second approximation is performed. On the other hand, the BPSK signal achieve a SR of 100% and a DA percentage of 0% at 2 dB and higher SNR.

Figures 2b and 2d depict the average error percentage of the amplitude, center frequency and bandwidth approximations made for the signals with each available SNR. As expected, the error is higher at low SNR, however it is always less than 10% for all the parameters in the BPSK signals with SNR values higher than 0 dB. In particular, BFSK signals with SNR values over -6 dB are approximated, on average, with an accuracy of 11.32% for amplitude, 5.11% for frequency and 2.90% for bandwidth. While BPSK signals with SNR values over 2 dB present accuracies of 0.78% for amplitude, 6.83% for frequency and 0.59% for bandwidth, on average.

On average, the amplitude is the parameter with the lowest error, while the frequency has the highest error. This is due to the asymmetry introduced by the noise in the FFT, i.e., if the signal FFT has a higher contribution of lower frequencies, the signal detection algorithm tends to converge at lower frequencies.

The signal detection algorithm was also tested for real signals transmitted by commercial modules in different frequencies and with BFSK modulation. The data were acquired using an Ettus N210 with a daughterboard UBX. The results for this experimental signals are presented in Table 3.

Band	SR [%]	DA [%]	AE [%]	FE [%]	BE [%]
VHF	87.50	100	47.65	0.66	49.99
UHF	97.50	100	0.60	0.73	23.92
S	99.50	100	15.00	0.33	6.52

Table 3: Results of the signal detection algorithm for real signals.

In the experimental results, the parameter estimation algorithm had the poorest results for VHF signals. This is due to the noisy environment where the tests were performed, as there were many signals interfering in that bandwidth.

5.2 Modulation classification

As discussed in Section 4, we use 8 different features to classify between BPSK, BFSK, CW and GMSK signals. For training the machine learning algorithms, we use the 80% of the previously mentioned datasets, while 20% are left for testing. In particular, 5 datasets of each SNR is used for testing, while 20 are used for training. Moreover, we define two classification problems: the base case and the binary case. Particularly, we use SVM for the binary case, while ANN with different numbers of neurons and layers are used for the base case. Specifically, we tested one and two layers with numbers of neurons varying from 2 to 20 in each layer. Figure 3 shows the classification accuracy obtained for each modulation type in both the base and the binary cases. Note that only the 4 ANN configurations with the best results are shown in the figure. The other results were omitted for clarity of the figure.

We observe that the best classification accuracies in the base case were obtained when only one layer was used for the ANN.

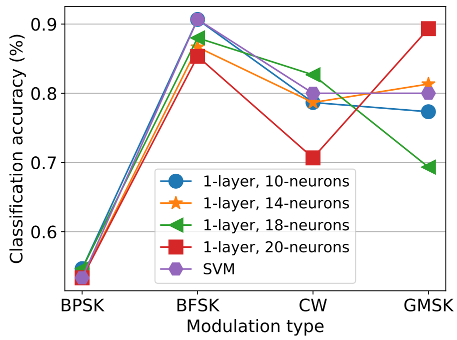


Figure 3: Classification accuracy of the tested methods.

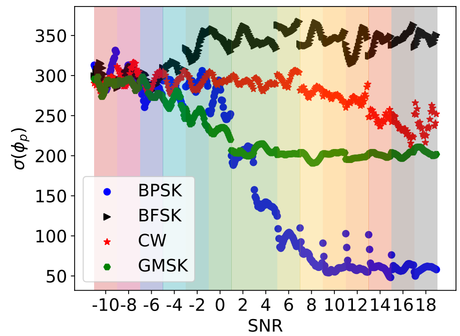


Figure 4: Standard deviation of instantaneous phase Vs SNR for all signals.

Generally, BPSK signals present the lowest accuracy, even though the most relevant feature is always the standard deviation of the instantaneous phase of the signals. This can be explained by the fact that BPSK signals highly overlap with other signal from different modulations, specially, at SNR values from -10 dB to 2 dB. This can be seen in Figure 4, where the standard deviation of the instantaneous phase of the signals is plotted against their SNR. In the figure, all the values inside a color bar, have the same SNR.

Although all approaches achieve similar results, we observe that the binary case with SVM has consistently high accuracies at all modulations. In contrast, the ANN approach with 1-layer and 20-neurons can classify GMSK signals with an accuracy of 89.3%, however, it achieves only 70.6% with CW signals.

Based on the presented results, SVM binary classifiers are better than the ANN to accurately classify the studied modulations. Moreover, since the implemented SVM use linear kernels for the binary classifications, the time complexity for testing new signals is negligible. In addition, all calculations can be made with a window of only 1024 samples, which makes the proposed approach suitable for real-time implementations.

6 CONCLUSIONS

This paper presented an algorithm for signal detection and modulation classification of signals typically used in amateur satellites. Specifically, our algorithm estimates the amplitude, center frequency and bandwidth with an error of less than 10% for signals with SNR higher than 0 dB. Moreover, our approach uses SVM to classify four different modulation types. By defining the classification problem as multiple binary classifiers in series, our approach achieves higher accuracies than ANN. As future work, we expect to embed the presented algorithm into an out-of-tree module in GNURadio.

7 ACKNOWLEDGMENTS

The authors would like to thank Ville Viinikka for his contributions to the literature review presented in this paper.

REFERENCES

[1] A. A. Boulogeorgos, N. D. Chatzidiamantis, and G. K. Karagiannidis. 2016. Energy Detection Spectrum Sensing Under RF Imperfections. *IEEE Transactions on Communications* 64, 7 (July 2016), 2754–2766. https://doi.org/10.1109/TCOMM.2016.2561294

[2] DK3WN. 2020. DK3WN SatBlog. Available in: http://www.dk3wn.info/p/?page_id=29535.

[3] J. Jagannath, N. Polosky, D. O'Connor, L. N. Theagarajan, B. Sheaffer, S. Foulke, and P. K. Varshney. 2018. Artificial Neural Network Based Automatic Modulation Classification over a Software Defined Radio Testbed. In *2018 IEEE International Conference on Communications (ICC)*. 1–6. https://doi.org/10.1109/ICC.2018.8422346

[4] G. Jajoo, Y. Kumar, and S. K. Yadav. 2019. Blind Signal PSK/QAM Recognition Using Clustering Analysis of Constellation Signature in Flat Fading Channel. *IEEE Communications Letters* 23, 10 (Oct 2019), 1853–1856. https://doi.org/10.1109/LCOMM.2019.2929127

[5] Xiaofan Li, Fangwei Dong, Sha Zhang, and Weibin Guo. 2019. A Survey on Deep Learning Techniques in Wireless Signal Recognition. *Wireless Communications and Mobile Computing* 2019 (2019), 5629572:1–5629572:12.

[6] Ryan A. Loomis, Karin I. Öberg, Sean M. Andrews, Catherine Walsh, Ian Czekala, Jane Huang, and Katherine A. Rosenfeld. 2018. Detecting Weak Spectral Lines in Interferometric Data through Matched Filtering. *The Astronomical Journal* 155, 4 (apr 2018), 182. https://doi.org/10.3847/1538-3881/aab604

[7] N2YO. 2020. Amateur radio satellites. Available in: https://www.n2yo.com/satellites/?c=18&srt=14&dir=1.

[8] Union of Concerned Scientists. 2019. UCS Satellite Database. Available in: https://www.ucsusa.org/resources/satellite-database.

[9] Tim O'Shea. 2017. *Open Radio ML Synthetic Benchmark Dataset*. https://github.com/radioML/dataset.

[10] T. J. O'Shea, T. Roy, and T. C. Clancy. 2018. Over-the-Air Deep Learning Based Radio Signal Classification. *IEEE Journal of Selected Topics in Signal Processing* 12, 1 (Feb 2018), 168–179. https://doi.org/10.1109/JSTSP.2018.2797022

[11] SatNOGS project. 2020. SatNOGS database. Available in: https://db.satnogs.org/.

[12] Hector Reyes, Sriram Subramaniam, Naima Kaabouch, and Wen Chen Hu. 2016. A spectrum sensing technique based on autocorrelation and Euclidean distance and its comparison with energy detection for cognitive radio networks. *Computers & Electrical Engineering* 52 (2016), 319 – 327. https://doi.org/10.1016/j.compeleceng.2015.05.015

[13] A. O. A. Salam, R. E. Sheriff, S. R. Al-Araji, K. Mezher, and Q. Nasir. 2015. Automatic Modulation Classification in Cognitive Radio Using Multiple Antennas and Maximum-Likelihood Techniques. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. 1–5. https://doi.org/10.1109/CIT/IUCC/DASC/PICOM.2015.3

[14] A. Smith, M. Evans, and J. Downey. 2017. Modulation classification of satellite communication signals using cumulants and neural networks. In *2017 Cognitive Communications for Aerospace Applications Workshop (CCAA)*. 1–8. https://doi.org/10.1109/CCAAW.2017.8001878

[15] Curtis M. Watson. 2010. *Signal Detection and Digital Modulation Classification-Based Spectrum Sensing for Cognitive Radio*. Ph.D. Dissertation. Northeastern University.