Evaluation of a new spectrum sensing technique for Internet of Things: An AI approach

Partemie M. Mutescu[1], Alexandru Lavric[1], Adrian I. Petrariu[1,2], Valentin Popa[1]
[1]Computers, Electronics and Automation Department, Stefan cel Mare University of Suceava, Romania
[2]MANSiD research center, Stefan cel Mare University of Suceava, Romania
marian.mutescu@usm.ro, lavric@usm.ro, apetrariu@usm.ro, valentin.popa@usm.ro

Abstract— **In the last decade we observed a great demand for wireless sensor applications as the connectivity of objects related to the Internet of Things concept increased. The growing number of wireless sensors leads to more spectrum demand and eventually to collisions due to overcrowding, causing a decrease in their performance level. Thus, to avoid collisions, detailed knowledge of the radio spectrum is required such as the degree of spectrum occupancy and the radio modulations used. This paper presents an analysis of the impact of different radio signal representations (I/Q coordinates, polar coordinates, and Fast Fourier Transform) on the performance level of machine learning algorithms in spectrum sensing classification. Our results shown that machine learning algorithms achieve a higher classification accuracy when the FFT representation of the radio signal is used, with a classification accuracy of 98.6%. When using the time series, the I/Q representation of the radio signal obtained an accuracy of 68.6% on the test dataset meanwhile the polar coordinates achieved an accuracy of 90%, respectively.**

Keywords—Spectrum sensing, Machine Learning, IoT, Modulation Classification, Fast Fourier Transform.

I. INTRODUCTION

The last decade has been marked by the evolution of IoT (Internet of Things) applications, mainly due to the rapid development of new wireless technologies. The vast majority of wireless communication protocols used in IoT applications operate in the unlicensed SRD radio spectrum (868 MHz) and ISM 2.4 GHz (e.g., LoRa [1], Sigfox [2]), as these frequency communication bands can be used by any device that respects local regulation rules. However, there are also IoT wireless communication protocols operating in the licensed spectrum that use the already deployed infrastructure to ensure a fast expansion of IoT services (e.g., NB-IoT [3] which uses the cellular communications infrastructure). The fast expansion of IoT services in both the unlicensed and licensed radio spectrum brings along a large number of IoT devices simultaneously connected. According to [4] there are currently 11.570 million IoT devices connected worldwide, and their number is expected to be at least double by the end of the decade. As the number of IoT devices is growing, the quality of service is significantly decreasing, due to radio spectrum overcrowding, data packet collisions, and problems regarding the coexistence of multiple wireless protocols.

Taking this into account it is mandatory to implement, develop and test spectrum sensing (SS) techniques that have the ability to solve the problems associated with the large-scale high-density wireless sensor networks, increasing the associated level of performance.

This paper is structured as follows: Section 1 presents an overview of SS and the implications of machine learning (ML) in SS techniques, Section 2 presents the workflow of the dataset creation process, as well as the main contribution of this paper which is represented by the evaluation of the ML algorithms performance when using different representations of radio signals: I/Q, Amplitude/Phase (A/P) and FFT as input data for the ML algorithms. Section 3 summarizes the results and the conclusions of this study.

II. SPECTRUM SENSING AND MACHINE LEARNING

A. Spectrum Sensing

SS is the main component of intelligent radio systems, also known as cognitive radio [5]. By sensing the radio spectrum, we can obtain valuable information regarding the occupancy level of the radio channels, the wireless communication protocol used as well as patterns regarding the activity and communication frequency of the wireless sensor network (WSN) on the communication channel. This information makes radio devices self-aware of the surrounding radio spectrum behavior and allows the WSN to coexist.

Current approaches to the problems associated with the large-scale high-density WSNs involve the use of energy detection methods [6] along with signal processing and artificial intelligence (AI) techniques [7] to detect unused frequency channels within the radio spectrum that can be used for communication, thus enhancing the performance of the wireless network. However, in order to gain a detailed knowledge of the radio spectrum, the detection of unused communication channels is considered to be insufficient. For this reason, in addition to the detection of unused communication channels, it is mandatory for us to detect and classify the radio modulation schemes associated with the communication protocols found within a wireless network.

978-1-6654-8162-5/22/$31.00 ©2022 IEEE

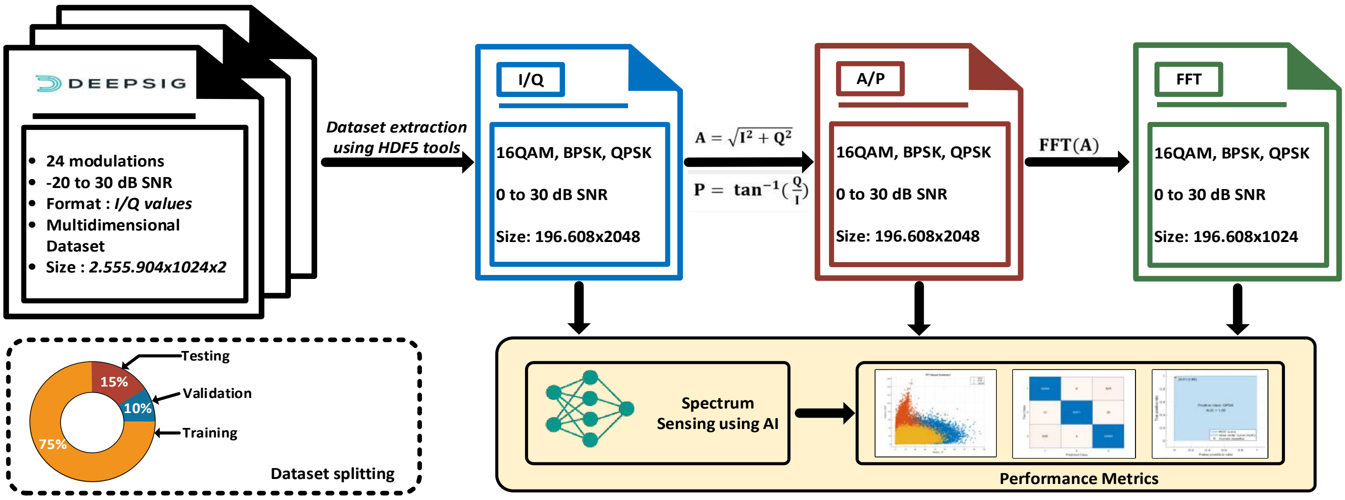

Figure 1. Study workflow

Practical implementations of SS techniques firstly involve the use of software defined radio (SDR) devices [8] for radio signal capturing [9]. The captured signals are then preprocessed and prepared to be passed through signal classification algorithms, the main tools used in signal classification being machine learning (ML) [10] and deep learning (DL) [11].

B. *Machine Learning and Spectrum Sensing*

ML techniques have been successfully assessed and implemented in SS applications for transmission detection [12], radio modulation recognition [13], and anomaly detection [14]. An ML algorithm learns how to perform a certain task related to a classification problem, based on a large dataset composed of training examples by detecting patterns in the data. The datasets used in SS can consist of parameters and features of the communication signal to be analyzed like I/Q samples, A/P samples, or FFT samples. Such a dataset was released by the DeepSig company [15] and is related to paper [11]. The dataset includes twenty-four different modulations (digital and analog) both from over-the-air radio signal captures and synthetically generated using SDR platforms. Each modulation has 4,096 instances at 26 different signal-to-noise (SNR) values ranging between -20 and +30 resulting in 106,496 instances for every modulation type. Each signal instance is composed of 1,024 samples consisting of I/Q value pairs. A detailed analysis of the dataset was presented in [16]. The dataset is encoded in the Hierarchical Data Format version 5 (HDF5) which is split into three groups X, Y, and Z. The X group is a three-dimensional space (2,555,904x1,024x2) containing the I/Q values of the signals, Y is a two-dimensional space (2,555,904x24) containing the modulation type indicator and Z contains the SNR value for each instance [16]. The dataset classes are stored in a separate file that is related to the Y group in the dataset.

Another approach in generating a dataset was described in [17]. This dataset introduces along with the radio signal samples additional parameters like base signal period, carrier offset, excess bandwidth, up sample factor, down sample factor, in-band SNR, and noise spectral density. In contrast to the DeepSig dataset, this one is smaller in size, with only 112,000 signal instances and 8 different radio modulation classes. Each signal instance is saved into a binary file containing a series of synthetically generated I/Q samples. The process of generating and validating a training dataset is one of the main challenges of the SS problem, as a synthetic dataset can't fully reproduce signal propagation effects (e.g., multipath fading). Taking this into account it is mandatory for the training dataset to contain training examples from over-the-air captures. The main purpose of this paper is to evaluate the impact of radio signal representation on the performance of ML algorithms in SS, as to our current knowledge such study hasn't been yet conducted.

III. SPECTRUM SENSING TECHNIQUES

A. *Preprocessing*

For our study, we chose the RadioML 2018.01A by DeepSig as a base dataset. Fig.1 shows the workflow of our study. Firstly, since the DeepSig dataset is large in size (24 radio modulations, 2,555,904 signal instances), we selected three radio modulations that are the closest related to IoT: 16QAM (used in 802.11a, and LTE-M communications), BPSK (used in Sigfox communications), QPSK (used in NB-IoT communications). As described before, the DeepSig dataset is encoded in an HDF5 format, which creates a problem that is related to the further processing of the dataset. Thus, in order to use this dataset, the data had to be extracted and preprocessed using HDF5 custom tools. Since the labels of the samples are described in a different file, further measures must be considered to ensure the synchronization of the extracted samples and, determine the corresponding positions in the dataset for all three modulations according to the additional dataset information provided by DeepSig and the dataset analysis presented in [16]. After determining the correct positions, we extracted 4,096 radio signal I/Q instances at 16 different SNR values ranging from 0 to 30 dB, for each one of

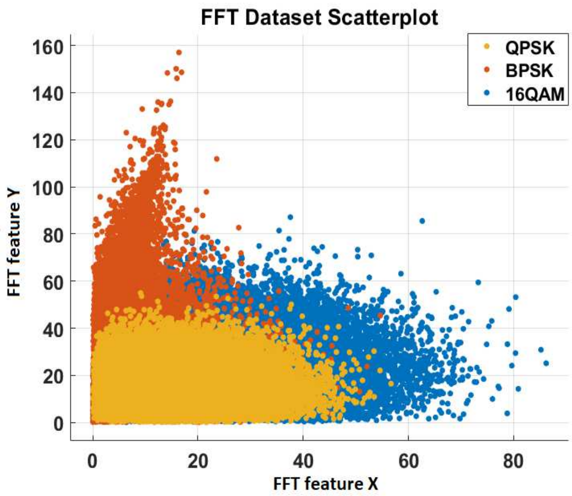

Figure 2. FFT dataset scatterplot

the three modulations, resulting in a first dataset consisting of 196,608 samples. Each sample consists of 1,024 pairs of I/Q coordinates arranged alternately. I/Q coordinates are the Cartesian representation of a signal's amplitude and phase. They are also known as the raw and unprocessed form of a radio signal, mainly because they can be retrieved directly from the output of a SDR receiver. Using the I/Q coordinates as input data for the tested ML classification algorithms would result in a drastic reduction of the complexity of a radio device with SS capabilities, as signals do not require any post-reception processing. Thus, the raw signals are applied to the AI algorithms. For the second dataset, we used the representation of the signal instances in polar coordinates (A/P). Both the amplitude and the phase of the signals were obtained by processing the I/Q pairs from the first dataset accordingly to the formulas presented in Fig. 1. The resulting dataset is composed of 196,608 radio signal instances each one represented by 1,024 A/P pairs. The third analyzed dataset was obtained by applying the FFT transformation on the amplitude time series of each signal instance, thus shifting the signals from the time domain to the frequency domain. Each dataset aims to reveal certain patterns in the signals to separate the radio modulation techniques and detect them with high accuracy.

B. AI Design and Development

The three resulting datasets were split into three parts: 150,423 samples were used for training; 16,713 samples were used for validation (10-fold cross-validation, approximative 10 %) and 29,472 samples were reserved for testing (approximative 15%). The datasets were then passed through the 31 ML algorithms available in Matlab [18] from which only three were selected (one for each dataset) based on the analysis of the resulting performance metrics: scatterplot of the analyzed features, confusion matrix which summarizes the classification results, ROC (Receiver Operating Characteristic) curve which measures the performance of the ML algorithm at different classification thresholds and AUC (Area Under the Curve). The I/Q dataset achieved a maximum accuracy of 68.4% on the validation dataset and 68.6% on the test dataset, using the Ensemble Bagged Trees algorithm. The A/P dataset showed a significant improvement in performance with an 89.6% and 90% accuracy on the validation dataset, respectively on the test dataset, using the Quadratic Discriminant algorithm. Further improvements have been achieved using the FFT dataset, with an accuracy of 98.6% on both the validation and test dataset, using the Quadratic SVM algorithm.

In Fig. 2 we can see a scatterplot of the FFT dataset for two distinctive features related to the amplitude of the signal at two different frequencies. From the scatterplot, we can see that the three different radio modulations are easily separable. Fig. 3 presents the confusion matrix and the ROC curve of the Quadratic SVM algorithm. The confusion matrix shows that the algorithm has a relatively low level of performance when distinguishing between the 16QAM and the QPSK modulations, as it misclassified 174 of the 16QAM signal instances as QPSK and 192 QPSK signal instances as 16QAM. From the ROC curve graph, we can see that the corresponding AUC is about 0.98. From the obtained results our study shows that ML algorithms perform better on frequency-series datasets than on time-series datasets when it comes to radio modulation classification problems. The best performance was obtained by the Quadratic SVM classification algorithm with an accuracy of 98.6% when classifying 16QAM, BPSK, and QPSK modulations. The sensitivity and specificity for each modulation class are presented in Table 1.

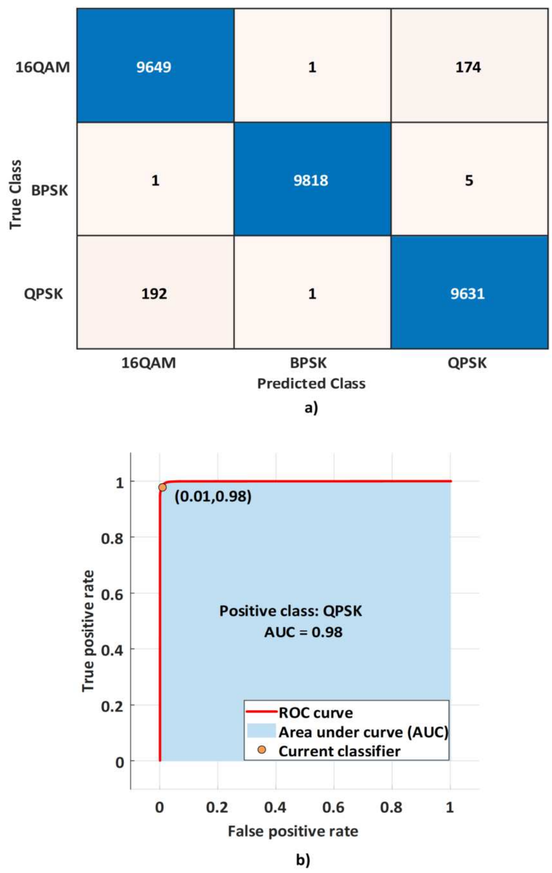

Figure 3. Quadratic SVM algorithm on FFT test dataset a) Confusion Matrix b) ROC curve for the QPSK modulation.

TABLE 1 SENSITIVITY AND SPECIFICITY PARAMETERS

	16QAM	BPSK	QPSK
Sensitivity	98.2%	99.93%	98.03%
Specificity	99.01%	99.98%	99.08%

IV. CONCLUSIONS

The recent evolution in both wireless technologies and IoT applications has led to an increase of wireless devices simultaneously connected, the radio spectrum resources being in high demand both in the licensed and unlicensed frequency band. SS techniques can help overcome these issues by giving both the network operators and the radio technology developers valuable information about the spectrum usage patterns, the radio modulations used in certain frequency bands, or the presence of anomalies within the radio spectrum. As modulation recognition represents a signal processing and a pattern recognition problem, machine learning approaches can be successfully integrated in SS techniques.

In this paper, we presented an analysis of the ML algorithms performance applied in radio modulation classification: 16QAM, BPSK, and QPSK. Different datasets consisting of different representations of the modulated radio signals (I/Q, A/P, and FFT) were used and evaluated using the developed AI approach. The results were ranked and evaluated by comparing the classification accuracies (validation and testing), confusion matrices, and ROC curves for each classification algorithm we tested.

The I/Q dataset achieved a maximum accuracy of 68.4% on the validation dataset and 68.6% on the test dataset using the Ensemble: Bagged Trees classification algorithm. The A/P dataset achieved an accuracy of 89.6% on the validation dataset and 90% on the test dataset using the Quadratic discriminant classification algorithm. The FFT dataset showed the highest performance with an accuracy of 98.6% for both validation and testing using the Quadratic SVM algorithm. From the obtained results, we concluded that machine learning algorithms achieve higher classification accuracy on frequency-series datasets rather than on time-series datasets. Taking this into account we believe that ML techniques for SS require further investigations, by gradually increasing the number and complexity of the modulation used as classes as well as testing and improving the robustness of ML algorithms to radio signal parameters variation.

ACKNOWLEDGMENT

This work was supported in part by the project Integrated Center for Research, Development and Innovation in Advanced Materials, Nanotechnologies, and Distributed Systems for Fabrication and Control (MANSiD), Sectoral Operational Program for Increase of the Economic Competitiveness co-funded from the European Regional Development Fund, under Contract 671/09.04.2015.

REFERENCES

[1] A. Lavric, A. I. Petrariu, E. Coca, and V. Popa, "LoRa Traffic Generator Based on Software Defined Radio Technology for LoRa Modulation Orthogonality Analysis: Empirical and Experimental Evaluation," *Sensors 2020, Vol. 20, Page 4123*, vol. 20, no. 15, p. 4123, Jul. 2020, doi: 10.3390/S20154123.

[2] A. Lavric, A. I. Petrariu, and V. Popa, "SigFox Communication Protocol: The New Era of IoT?," *2019 International Conference on Sensing and Instrumentation in IoT Era, ISSI 2019*, Aug. 2019, doi: 10.1109/ISSI47111.2019.9043727.

[3] R. S. Sinha, Y. Wei, and S. H. Hwang, "A survey on LPWA technology: LoRa and NB-IoT," *ICT Express*, vol. 3, no. 1, pp. 14–21, Mar. 2017, doi: 10.1016/J.ICTE.2017.03.004.

[4] "• IoT connected devices by region 2030 | Statista." https://www.statista.com/statistics/1194677/iot-connected-devices-regionally/ (accessed Mar. 29, 2022).

[5] J. Mitola and G. Q. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, no. 4, pp. 13–18, Aug. 1999, doi: 10.1109/98.788210.

[6] N. N. Dao, W. Na, A. T. Tran, D. N. Nguyen, and S. Cho, "Energy-Efficient Spectrum Sensing for IoT Devices," *IEEE Systems Journal*, vol. 15, no. 1, pp. 1077–1085, Mar. 2021, doi: 10.1109/JSYST.2020.2986030.

[7] B. al Homssi, A. Al-Hourani, Z. Krusevac, and W. S. T. Rowe, "Machine Learning Framework for Sensing and Modeling Interference in IoT Frequency Bands," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4461–4471, Mar. 2021, doi: 10.1109/JIOT.2020.3026819.

[8] J. Mitola, "Software radios-survey, critical evaluation and future directions," *Proceedings - National Telesystems Conference, NTC 1992*, 1992, doi: 10.1109/NTC.1992.267870.

[9] A. Nafkha, M. Naoues, K. Cichon, and A. Kliks, "Experimental spectrum sensing measurements using USRP Software Radio platform and GNU-radio," *Proceedings of the 2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CROWNCOM 2014*, pp. 429–434, 2014, doi: 10.4108/ICST.CROWNCOM.2014.255415.

[10] E. Almazrouei, G. Gianini, N. Almoosa, and E. Damiani, "What can Machine Learning do for Radio Spectrum Management?," *Q2SWinet 2020 - Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks*, pp. 15–21, Nov. 2020, doi: 10.1145/3416013.3426443.

[11] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-Air Deep Learning Based Radio Signal Classification," *IEEE Journal on Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 168–179, Feb. 2018, doi: 10.1109/JSTSP.2018.2797022.

[12] T. Gale, T. Šolc, R. A. Moşoi, M. Mohorčič, and C. Fortuna, "Automatic Detection of Wireless Transmissions," *IEEE Access*, vol. 8, pp. 24370–24384, 2020, doi: 10.1109/ACCESS.2020.2970840.

[13] B. Jdid, K. Hassan, I. Dayoub, W. H. Lim, and M. Mokayef, "Machine Learning Based Automatic Modulation Recognition for Wireless Communications: A Comprehensive Survey," *IEEE Access*, vol. 9, pp. 57851–57873, 2021, doi: 10.1109/ACCESS.2021.3071801.

[14] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, "SAIFE: Unsupervised Wireless Spectrum Anomaly Detection with Interpretable Features," *2018 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN 2018*, Jan. 2019, doi: 10.1109/DYSPAN.2018.8610471.

[15] "Reinventing Wireless with Deep Learning | DeepSig." https://www.deepsig.ai/ (accessed Mar. 18, 2022).

[16] "DeepSig's 2018 Data Set: 2018.01.OSC.0001_1024x2M.h5.tar.gz – Cyclostationary Signal Processing." https://cyclostationary.blog/2020/09/24/deepsigs-2018-data-set-2018-01-osc-0001_1024x2m-h5-tar-gz/ (accessed Mar. 18, 2022).

[17] "Data Set for the Machine-Learning Challenge – Cyclostationary Signal Processing." https://cyclostationary.blog/2019/02/15/data-set-for-the-machine-learning-challenge/ (accessed Mar. 24, 2022).

[18] "MATLAB - MathWorks - MATLAB & Simulink." https://www.mathworks.com/products/matlab.html (accessed Mar. 24, 2022)