Unsupervised Wireless Spectrum Anomaly Detection With Interpretable Features

Sreeraj Rajendran, Wannes Meert, Vincent Lenders, *Member, IEEE*, and Sofie Pollin, *Senior Member, IEEE*

***Abstract*—Detecting anomalous behavior in wireless spectrum is a demanding task due to the sheer complexity of the electromagnetic spectrum use. Wireless spectrum anomalies can take a wide range of forms from the presence of an unwanted signal in a licensed band to the absence of an expected signal, which makes manual labeling of anomalies difficult and suboptimal. We present, spectrum anomaly detector with interpretable features (SAIFE), an adversarial autoencoder (AAE)-based anomaly detector for wireless spectrum anomaly detection using power spectral density (PSD) data. This model achieves an average anomaly detection accuracy above 80% at a constant false alarm rate of 1% along with anomaly localization in an unsupervised setting. In addition, we investigate the model's capabilities to learn interpretable features, such as signal bandwidth, class, and center frequency in a semi-supervised fashion. Along with anomaly detection the model exhibits promising results for lossy PSD data compression up to 120× and semi-supervised signal classification accuracy close to 100% on three datasets just using 20% labeled samples. Finally, the model is tested on data from one of the distributed electrosense sensors over a long term of 500 h showing its anomaly detection capabilities.**

***Index Terms*—Deep learning, spectrum monitoring, anomaly detection.**

I. Introduction

THE NEW generation of wireless technologies is promising improved throughput, latency and reliability enabling the creation of novel applications. The fifth generation wireless deployments will be very heterogeneous ranging from millimeter wave communications to massive MIMO and LoRa/Sigfox deployments for line of sight (LOS), medium and long range communication systems respectively. Such dense and heterogeneous deployment makes the enforcement and management of the wireless spectrum usage difficult. In addition, manual spectrum management is inefficient and can only deal with a limited number of anomalies and measurement locations. Complex spectrum regulations across frequency bands in various countries along with illegal interference worsen the problem. Automated spectrum monitoring solutions covering frequency, time and space dimensions are becoming more crucial than ever before.

Unlike other sensing contexts such as air quality, temperature or city traffic monitoring, wireless spectrum monitoring on a large scale raises many unique problems ranging from the data costs associated with the sheer volume of sensed spectrum information to sensor quality and data privacy issues. These wide ranging infrastructure problems were systematically analyzed and partially solved by the Electrosense [1] platform. Electrosense is interdisciplinary and combines the power of crowdsourcing with Big data to solve the wireless spectrum monitoring problem. The sensing devices are low cost Software Defined Radio (SDR) dongles connected to embedded devices like a Raspberry Pi or high end SDR devices connected through a personal computer. Through Electrosense, an Open Spectrum Data as a Service (OSDaaS) model was introduced to address the usability of the spectrum data for a wide range of stakeholders including wireless operators, spectrum enforcement agencies, military and generic users.

In addition to the sensor infrastructure problems that were tackled in Electrosense, various algorithmic challenges still need to be addressed to provide advanced spectrum utilization awareness. The central coup to achieve this vision is a wireless spectrum anomaly detector which can continuously monitor the spectrum and detect unexpected behavior. Furthermore, in addition to the detection of anomalies, it is important to understand the cause of an anomaly. This ranges from an unexpected transmission in the analyzed band that can be classified [2], to absence of an expected signal. Wireless anomaly detection to some extent has been addressed in wireless sensor networks in the past [3]–[5]. These techniques make use of derived expert features from very low rate sensor data such as temperature and pressure instead of high volume radio physical layer data as is our interest. An anomaly detector for Dynamic Spectrum Access (DSA) is presented in [6], where distributed power measurements via cooperative sensing are used for anomaly detection. The proposed detector is limited to authorized user anomaly detection only, for the specific case of DSA. Similarly [7] makes use of Hidden Markov Models (HMM) on spectral amplitude probabilities that can detect interference on the channel of interest again in the DSA domain.

Recently in [8], the authors presented a recurrent anomaly detector based on predictive modeling of raw In-phase and

Manuscript received December 6, 2018; revised March 1, 2019; accepted April 8, 2019. Date of publication April 16, 2019; date of current version September 9, 2019. This research was sponsored in part by the department of Science and Technology, armasuisse and the North Atlantic Treaty Organization (NATO) Science for Peace and Security Programme under grant G5461. The associate editor coordinating the review of this paper and approving it for publication was L. Duan. *(Corresponding author: Sreeraj Rajendran.)*

S. Rajendran and S. Pollin are with the Department ESAT, KU Leuven, 3001 Leuven, Belgium (e-mail: sreeraj.rajendran@esat.kuleuven.be; sofie.pollin@esat.kuleuven.be).

W. Meert is with the Department of Computer Science, KU Leuven, 3001 Leuven, Belgium (e-mail: wannes.meert@cs.kuleuven.be).

V. Lenders is with the Department of Science and Technology, armasuisse, 3602 Thun, Switzerland (e-mail: vincent.lenders@armasuisse.ch).

Digital Object Identifier 10.1109/TCCN.2019.2911524

quadrature phase (IQ) data. The authors used a Long Short Term Memory (LSTM) model for predicting the next 4 IQ samples from the past 32 samples and an anomaly is detected based on the prediction error. Even though this model works on raw physical layer data which requires no expert feature extraction, it is still not sufficiently automated and generic for practical anomaly detection. First, different copies of the same model need to be trained for different wireless bands such that the model is able to predict anomalies specific to the band of interest. For instance, an LTE signal in the FM broadcast band is definitely an anomaly thus preventing a single model to be trained on both bands. Second, the model does not extract any interpretable features to understand the cause of the anomaly. In [9], the authors extend this prediction idea on spectrograms and test the model on some synthetic anomalies. A reconstruction based anomaly detector based on vanilla deep autoencoders is presented in [10]. This model lacks interpretable feature extraction properties like class labels which implies the need for training multiple copies of the same model on different bands.

In this paper we argue that, reconstruction based anomaly detection could be superior to prediction based techniques as prediction is a tougher problem than reconstruction in complex time series datasets. For instance, while digitally modulating signals the basic assumption is each constellation point is selected with equal probability to maximize the information transfer which makes the prediction of the future symbols difficult. On the other hand reconstruction of input data from compressed features is an easier problem if the model can efficiently capture the complex data distributions. In addition, the time-varying random wireless channel makes prediction of future samples difficult outside the channel coherence time.

We propose SAIFE, an AAE based model which fills the shortcomings of these state-of-the-art (SoA) models. First, we show that a *single model can be trained over multiple bands* in an unsupervised fashion avoiding the need for multiple copies of models on various bands. Second, the same model can be *trained in a semi-supervised fashion for extracting interpretable features* such as signal bandwidth and position. Third, the reconstructed signal from the proposed model can be used for *localizing anomalies* in the wireless spectrum. Furthermore we explore various other advantages of the model such as *wireless data compression* and *signal classification* which are significant contributions in contrast to the SoA models [8]–[10].

The rest of the paper is organized as follows. The anomaly detection problem is clearly stated in Section II. Section III explains the AAE model used for anomaly detection along with the implementation details. The dataset and the parameters used for training are presented in Section IV. Section V details the performance results and discusses the advantages of the proposed model. Section VI explores the signal compression and classification features of the model. Conclusions and future work are presented in Section VII.

II. Problem Definition

Given: Let X_S be the source time-series data, where $\mathbf{x} \in X_S$ could be either a complex IQ vector or a frequency-based PSD vector from any wireless frequency band. The dataset X_S contains wireless signals that are assumed to be normal behavior. Thus the probability of anomalous behavior in this source dataset is assumed to be low. The superset $X_S = X_{S0} \cup X_{S1} \cdots \cup X_{Sn}$ contains signals from various frequency bands.

Goal: A model that learns the source data distribution $p(X_S)$ and detects when a target vector's distribution deviates from the source data distribution. For each target vector $\mathbf{x} \in X_T$, X_T being the test dataset, the model should infer whether the vector is normal (H_0) or anomalous (H_a), where H_0 and H_a are hypothesis listed below.

- H_0: Sample data comes from $p(X_S)$
- H_a: Sample data does not come from $p(X_S)$

A signal type of $\mathbf{x} \in X_{Sl}$ from frequency band l occurring in a band k where we are expecting X_{Sk} is also an anomalous behavior which demands the model to capture class labels for fine grained anomaly detection.

Assumptions:

1) The probability of anomalous behavior in the source dataset is very low.
2) No explicit anomaly labeling is done on the source and target dataset.
3) No expert feature extraction is performed before feeding data to the model.

III. Models

We leverage the recent advances in generative modeling using neural networks which are trained through backpropagation directly from data [11]–[14]. The key insight of these previous work is to bring the higher dimensional input data to some lower dimensional latent space ($\mathbf{Z}$), whose prior distributions can be specified. This latent space which captures relevant features or settings can be then used to reconstruct the actual input data, ideally with minimal reconstruction loss. A basic introduction to some of the recent SoA generative models are covered in the following subsections.

A. Autoencoder and Variational Autoencoder (VAE)

A traditional autoencoder, as shown in Figure 1, is a neural network that consists of an encoder (E) and a decoder (D). The encoder and decoder are trained to reduce the reconstruction loss. This entire network basically performs a non-linear dimensionality reduction optimizing the encoder and decoder parameters (θ and ϕ), the neural network weights, to achieve minimum reconstruction loss such as minimum squared error as given below

$$\theta, \phi = \arg\min_{\theta,\phi} ||\mathbf{x} - \hat{\mathbf{x}}||^2 \tag{1}$$

A VAE [11] also makes use of an its encoder-decoder structure. VAEs encode the input data vector to a vector $\mathbf{z}$ in the latent space $\mathbf{Z}$ whose priors can be imposed by using a Kullback-Leibler (KL) divergence penalty. VAE optimizes the

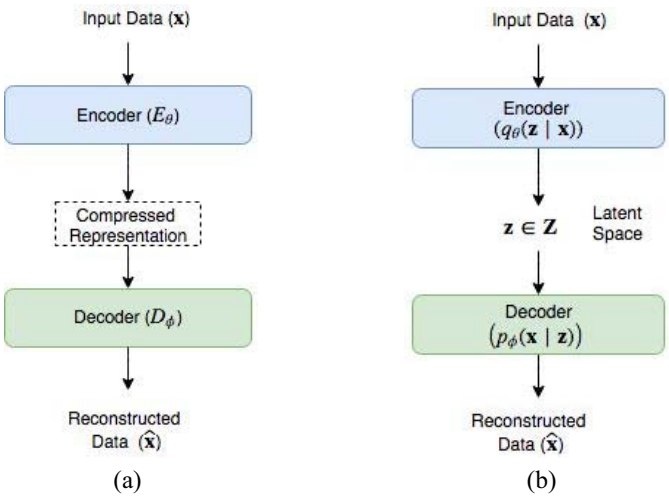

Fig. 1. (a) Encoder decoder structure of an unsupervised vanilla autoencoder model, (b) Stochastic variant of the autoencoder where the internal representations are probability distributions in general. (a) Vanilla autoencoder. (b) Variational autoencoder.

network parameters θ and ϕ to minimize the following upper-bound on the negative log-likelihood of $\mathbf{x}$, where p_{data} is the distribution of the data $\mathbf{x}$:

$$\begin{aligned}&\mathbb{E}_{\mathbf{x}\sim p_{data}}\big[-log\ p_\phi(\mathbf{x})\big]\\&< \mathbb{E}_{\mathbf{x}\sim p_{data}}\Big[\mathbb{E}_{\mathbf{z}\sim q_\theta(\mathbf{z}|\mathbf{x})}\big[-log\big(p_\phi(\mathbf{x}|\mathbf{z})\big)\big]\Big]\\&\quad+\mathbb{E}_{\mathbf{x}\sim p_{data}}\big[KL\big(q_\theta(\mathbf{z}|\mathbf{x})||p_\phi(\mathbf{z})\big)\big]\end{aligned}\tag{2}$$

Thus VAE optimize the reconstruction loss (first term) similar to a standard autoencoder but adds regularization terms (second term: KL divergence or cross-entropy term) which helps it to learn a latent representation that is consistent with the defined prior $p_\phi(\mathbf{z})$.

B. Adversarial Autoencoder (AAE)

Adversarial autoencoders [12] make use of the recent advances in generative modeling [13] to replace the KL divergence in VAEs with adversarial training that encourages the decoder to map the imposed prior to the data distribution. Thus AAE provides two major advantages over VAE: (i) the model ensures that the decoder will generate meaningful samples if we sample from any part of the prior space and (ii) as the aggregate posterior matches the prior distribution, variations of these distributions can be used for detecting unknown data inputs which is very useful for applications such as anomaly detection. In addition, AAE provides a flexible and robust architecture for semi-supervised learning and data visualization.

C. SAIFE Description

We make use of a deep learning model based on AAE to enable all the requirements mentioned in the problem definition as shown in Figure 2. An LSTM layer with 512 cells is used as the encoder for extracting interpretable features while a Convolutional Neural Network (CNN) based decoder is employed for reconstructing the input data from the extracted

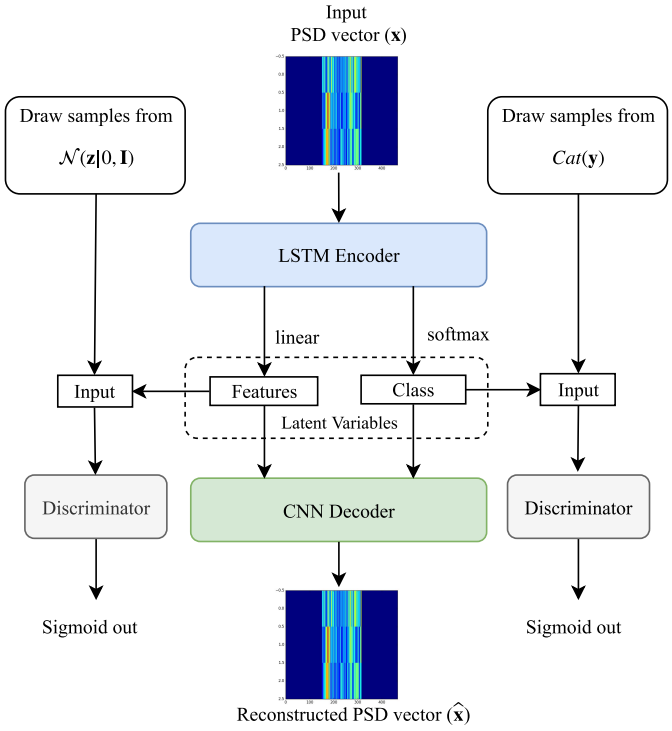

Fig. 2. Model architecture for anomaly detection.

features. The AAE architecture is trained in a semi-supervised fashion for making the features more interpretable while the reconstruction is fully unsupervised. Two layer feed forward networks with 256 cells and relu activations are employed in both discriminators. The LSTM output is fed through a softmax layer for signal classification and a linear layer for extracting the latent features.

The discriminators (D_s) are neural networks that evaluate the probability that the latent code $\mathbf{z}$ is from the prior distribution $p(\mathbf{z})$ that we are trying to impose rather than a sample from the output of the encoder (E) model. The discriminator receives $\mathbf{z}$ from both the encoder and the prior distribution and is trained to distinguish between them. The encoder is trained to confuse the discriminators into believing that the samples it generates are from the prior distribution. Thus the encoder is trained to reach the solution by optimizing both networks by playing a min-max adversarial game which is expressed in [13] as

$$\begin{aligned}\min_E \max_{D_s}\ &\mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}[log(D_s(\mathbf{z}))]\\&+\mathbb{E}_{\mathbf{x}\sim p_{data}}[log\ (1-D_s(E(\mathbf{x})))]\end{aligned}\tag{3}$$

Generative models try to model the underlying distributions of the input data, the latent variables, which are further used for data reconstruction. In SAIFE, the input PSD data is assumed to be generated by the latent *Class* variable which comes from a Categorical distribution with number of categories $k = number\ of\ frequency\ bands$ and the continuous latent *Features* from a Gaussian distribution of zero mean and unit variance; $p(\mathbf{y}) = Cat(\mathbf{y})$ and $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|0,\mathbf{I})$.

TABLE I
SYNTHETIC SIGNAL DATASET PARAMETERS

Type	single-cont, single-rshort, mult-cont, dethop
Input frame size	6x64
SNR Range	5dB to +20dB
Number of training samples	6000 vectors
Number of test samples	6000 vectors

TABLE II
SYNTHETIC ANOMALY DATASET PARAMETERS

Type	scont, randpulses, wpulse, oclass
Input frame size	6x64
SNR Range	-20dB to +20dB
Number of training samples	6000 vectors
Number of test samples	6000 vectors

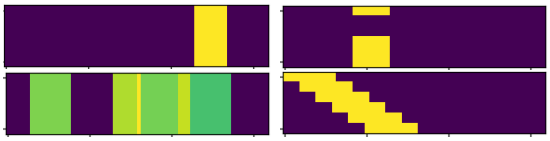

Fig. 3. Sample signals *single-cont*, *single-rshort*, *mult-cont* and *dethop* from synthetic signal dataset (time on y-axis and frequency on x-axis).

TABLE III
SDR AND ELECTROSENSE DATASET FREQUENCY BANDS

Dataset and bands	Frequencies (MHz)
SDR dataset, 0-11	80-107, 109-115.5, 117-140, 166-172, 196-208, 212.5-217.5, 220-227.5, 388-396, 422-427, 640-660, 790-800, 920-960.
Electrosense dataset, 0-6	86-108, 192-197, 790-801, 801-810, 811-821, 933-935, 955-960.

D. Implementation Details

The model is implemented using TensorFlow [15], a data flow graph based numerical computation library from Google. Python and C++ bindings of Tensorflow makes the usage of the final trained model easily portable to host based SDR frameworks like GNU Radio [16]. The trained model can be easily imported as a block in GNU Radio which can be readily used in practice with any supported hardware front-end. These models can be quantized and deployed on FPGAs or embedded CPUs close to the radio frontend for improving processing speeds [2], [17]. In addition, the availability of new machine learning accelerators such as Intel neural compute stick combined with low-cost single-board computers such as Raspberry Pi enable wide deployment and realtime inference of deep learning models easier.

IV. DATASETS AND MODEL TRAINING

We use three spectrum datasets along with one synthetic anomaly set to evaluate the performance of the used model. A synthetic spectrum dataset is necessary to understand the performance of the model in a controlled environment. The synthetic dataset consists of four different signal types with signal parameters as reported in Table I. The signals being (i) *single-cont*: single continuous signal with random bandwidth, signal-to-noise ratio (SNR) and center frequency, (ii) *single-rshort*: pulsed signals in time with similar parameters as *single-cont*, (iii) *mult-cont*: multiple continuous signals with possible overlap and (iv) *dethop*: random bandwidth and SNR signals with deterministic shifts/hops in frequency as depicted in Figure 3. Similarly, four synthetic signals (i) *scont*: same as single-cont, (ii) *randpulses*: random pulsed transmissions on the given band, (iii) *wpulse*: pulsed wideband signals covering the entire frequency, (iv) *oclass*: signals from other classes in synthetic dataset are used as anomalies.

In addition to the synthetic dataset we validate using two real wireless datasets. The first is a SDR dataset collected using a HackRF SDR from two different cities in Belgium covering frequencies from 10 MHz to 3 GHz. HackRF with its firmware sweep mode can scan the spectrum at up to 8 GHz per second, which allows scanning of 0-6 GHz under a second. Twelve frequency bands are selected from these spectrum scans, continuous in time covering various audio and video broadcast, GSM and LTE bands with a frequency resolution of 100 KHz whose frequency ranges are listed in Table III. The second dataset consists of PSD sensor data from multiple Electrosense sensors deployed all over Europe retrieved through the open API [1] with 7 selected frequency bands as listed in Table III.

A. Model Training

All the datasets mentioned in the previous section are split into two subsets, a training and a testing subset, with equal number of vectors. A seed is used to generate random mutually exclusive array indices, which are then used to split the data into two ascertaining the training and testing sets are entirely different. The train and test datasets for performance analysis on the SDR and Electrosense datasets are selected from different, non-adjacent, time periods to make sure that there are no identical points in the training and test dataset. The model is trained in an unsupervised fashion to reduce the mean squared error between the input and decoder output and a semi-supervised fashion to learn the continuous features and class labels. The adversarial networks as well as the autoencoder are trained in three phases: the reconstruction, regularization and semi-supervised phase as mentioned in [12]. The Adam optimizer [18], a first-order gradient based optimizer, with a learning rate of 0.001 is used for training in all the phases. In the semi-supervised phase the model is trained to learn the class, position and bandwidth of the input signal by training it on 20% of the labeled samples from the training set. The model is trained for 500 epochs which takes around one hour of training time on a x86 PC with Nvidia GeForce GTX 980 Ti graphics card. Once trained, the profiled model inference time for 500 input spectrograms is 0.0456 seconds on the same hardware resulting in a processing time of around 0.91 microseconds per input vector.

V. ANOMALY DETECTION

Once the training process is complete, the model weights are frozen and new input data is fed to the model. As mentioned in the model architecture section, anomalies are detected primarily based on the reconstruction error of the

model. In addition to the reconstruction error, the classification error and the discriminator loss are also used for detecting anomalous behaviors.

A. Detection Scores

Three scores are used to detect whether the input data frame is anomalous or not. They are

1) *Reconstruction Loss:* This error measures the similarity between the input data and the reconstructed data defined as $R_l = \sum_{i=0}^{N} |\mathbf{x} - \hat{\mathbf{x}}|$ where $\mathbf{x}$ is the frame input, $\hat{\mathbf{x}} = D(\mathbf{z})$ is the decoder frame output and N is the number of data points in the frame.
2) *Discriminator Loss:* The discriminator in the AAE model is trained to distinguish between the samples from the prior distribution and the samples generated by the encoder. We use the same discrimination loss used during the training process which is defined as $D_l = \sigma(\mathbf{z}, 1)$ where σ is the sigmoid cross entropy. The loss from both continuous (D_{lcont}) and categorical (D_{lcat}) discriminators are used for computing the final anomaly score.
3) *Classification Error:* The class labels predicted by the encoder is cross checked with the original band of interest for detecting the presence of other known but unexpected signals in a selected frequency band.

A simple n-sigma threshold is employed on the reconstruction and discriminator loss based on the mean and standard deviation values from the training data. An input data frame is classified as anomalous if A_{score} is *True*:

$$\begin{aligned} A_{score} = &\left(R_l > \left(\mu_{R_{lt}} + n * \sigma_{R_{lt}}\right)\right) \\ &\vee\left(\left(\mu_{D_{ltcont}} - n * \sigma_{D_{ltcont}}\right) > D_{lcont}\right. \\ &\qquad \left. > \left(\mu_{D_{ltcont}} + n * \sigma_{D_{ltcont}}\right)\right) \\ &\vee\left(\left(\mu_{D_{ltcat}} - n * \sigma_{D_{ltcat}}\right)\right. \\ &\qquad \left. > D_{lcat} > \left(\mu_{D_{ltcat}} + n * \sigma_{D_{ltcat}}\right)\right) \\ &\vee\left(Class_{Encoder}! = Class_{input}\right) \end{aligned} \tag{4}$$

The threshold value n is selected empirically based on the expected true positive rate and false detection rate. From the probability distributions of *dethop* signal and *dethop* signal with *scont* anomaly shown in Figure 4, it can be clearly noticed that the reconstruction loss along with class labels plays a major for anomaly detection.

B. Performance Comparisons

To evaluate the performance of SAIFE, the anomaly detection performance is compared against various SoA algorithms such as One class Support Vector Machine (OSVM), Isolation Forest (IFO) [19], Lightweight on-line detector of anomalies (LODA) [20] and Robust Covariance (RCOV) [21]. The average anomaly detection accuracy of these algorithms over different frequency bands are plotted in Figure 5. On an average SAIFE performs better than all other algorithms for all anomalies on all synthetic frequency bands. *Oclass* anomaly performance is quite good when compared to other algorithms as SAIFE performs explicit frequency band classification as one of the features. Detailed detection accuracies for various bands are shown in Figures 6, 7 and 8 for reference. Figure 10 shows the Receiver operating characteristic (ROC) curves

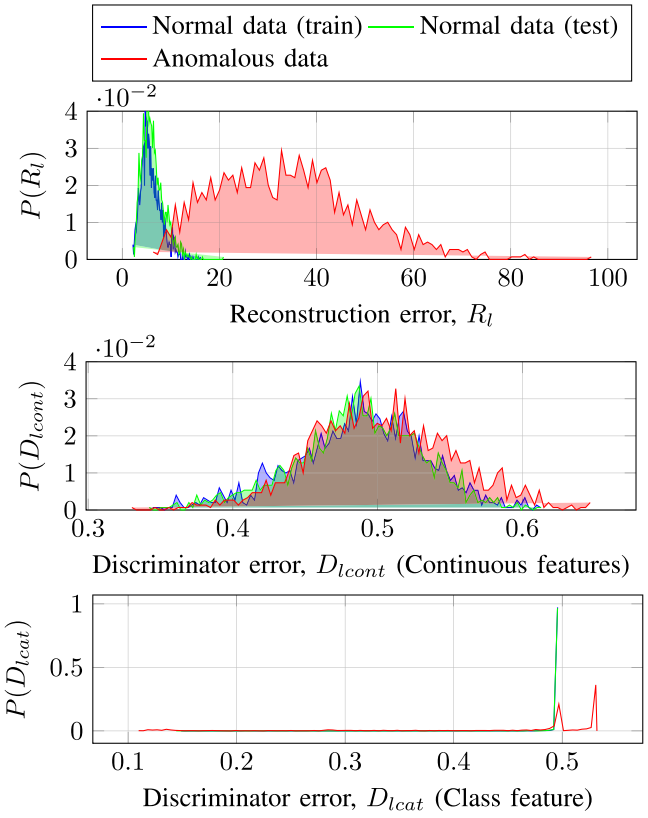

Fig. 4. Probability density functions of reconstruction error, continuous discriminator error and categorical discriminator error for *dethop* signal and *dethop* signal with *scont* anomaly.

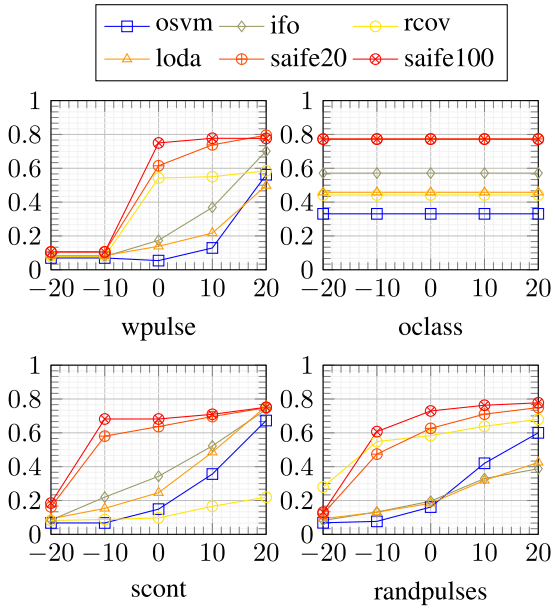

Fig. 5. Anomaly detection accuracies for different anomalies with a constant false alarm rate of 10% averaged over four different frequency bands. For *oclass* anomaly, anomaly vectors are randomly selected from other classes without specific SNR based evaluation resulting in one detection accuracy value (plotted as a line for uniformity). Anomaly SNR on the x-axis and detection accuracy on the y-axis.

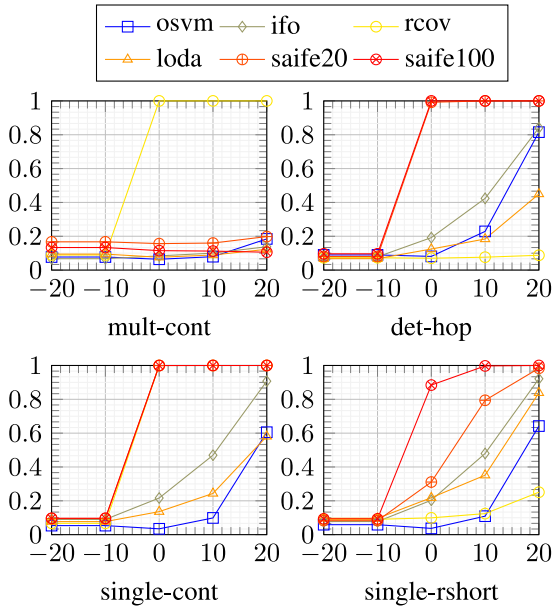

Fig. 6. Anomaly detection accuracies for *wpulse* anomaly with a constant false alarm rate of 10% on four different frequency bands. Anomaly SNR on the x-axis and detection accuracy on the y-axis.

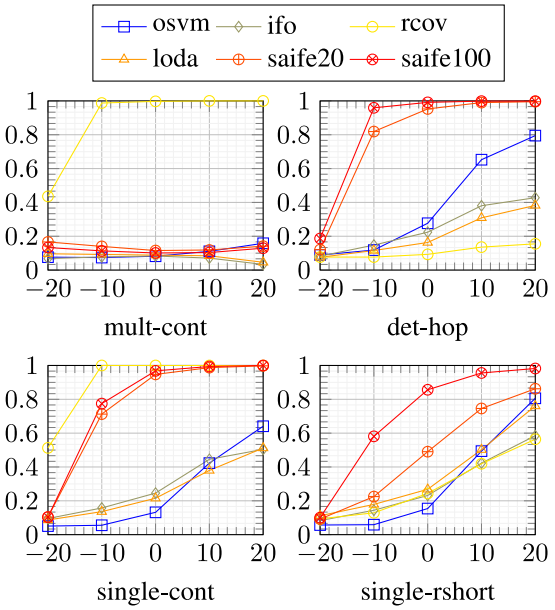

Fig. 8. Anomaly detection accuracies for *randpulses* anomaly with a constant false alarm rate of 10% on four different frequency bands. Anomaly SNR on the x-axis and detection accuracy on the y-axis.

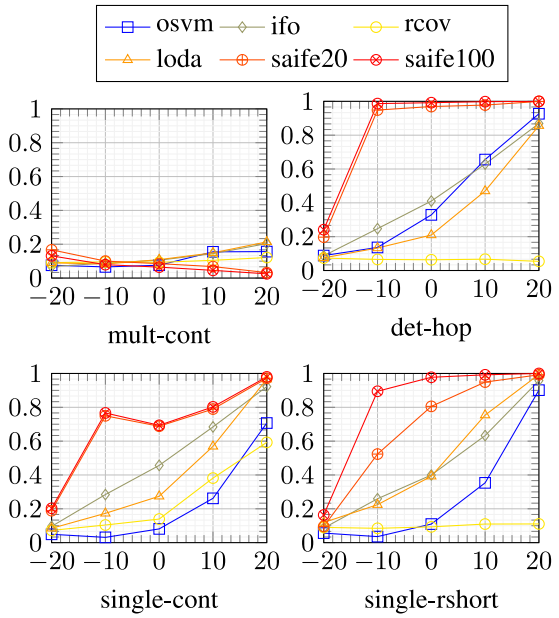

Fig. 7. Anomaly detection accuracies for *scont* anomaly with a constant false alarm rate of 10% on four different frequency bands. Anomaly SNR on the x-axis and detection accuracy on the y-axis.

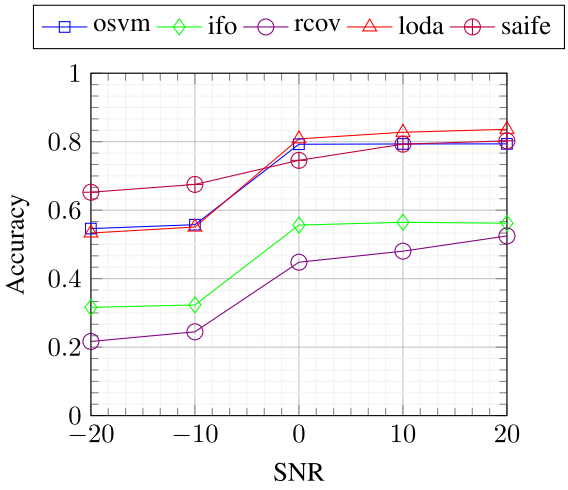

Fig. 9. Anomaly detection accuracies for different algorithms with a constant false alarm rate of 1% averaged over all frequency bands and anomalies. Anomaly SNR on the x-axis and detection accuracy on the y-axis.

for all anomalies on the *det-hop* channel for all algorithms. Anomaly signals similar to the original signals are intensionally selected to thoroughly analyze the detection capabilities of the model. For instance, from Figure 7, it can be seen that detection of *scont* anomaly is difficult in *mult-cont* band as another continuous signal is not an anomalous behavior in the multiple continuous signal band. Similarly detection of *wpulse* works well only on SNRs above 0 dB as the signal is only visible above the noise floor above 0 dB. Improving the number of features of SAIFE from 20 to 100 can also help to improve the detection accuracy to some extent as shown in Figure 5. More detailed analysis on optimal number of features can be found in Section VI.

These experiments are repeated on the SDR dataset and the results are plotted in Figure 11. Only the two best and worst performing frequency bands for different anomalies based on the Area Under Curve (AUC) for the lowest anomaly SNR of -20dB are shown due to space limitations, as there are 12 frequency bands in the SDR dataset. Results similar to the synthetic dataset can be noticed in the real capture SDR dataset also. Detecting *scont* and *randpulses* anomalies in frequency band 0 (80-107 MHz) is very difficult as the selected band is very wide and it contains strong FM broadcast stations. Similar results can be noticed in the other worst performing

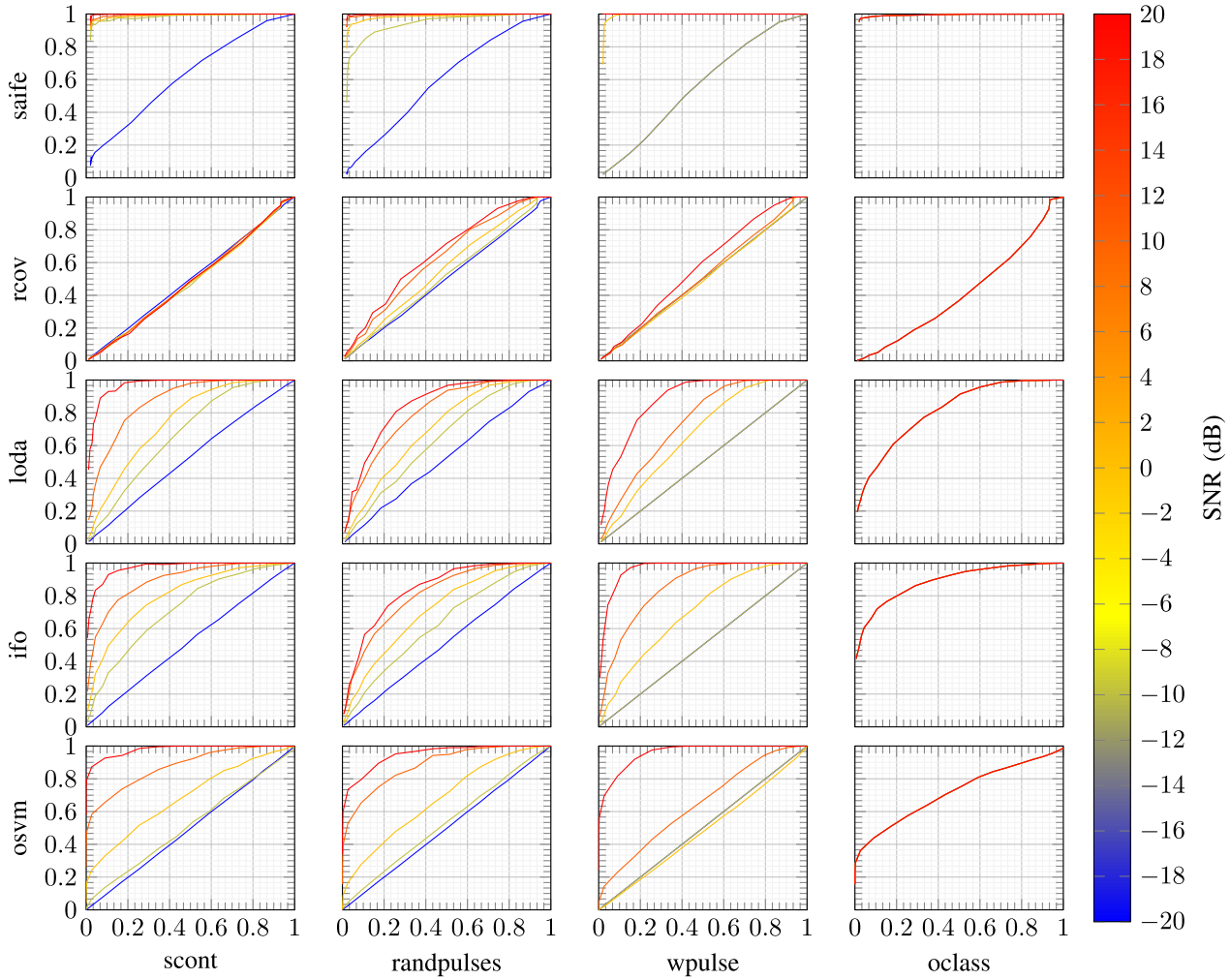

Fig. 10. ROC curves for different detection algorithms on *det-hop* synthetic band for various anomalies.

band 11 (920-960 MHz) which contains GSM signal transmissions that includes both continuous and hopping transmissions. This shows the pressing need to split the 40 MHz bandwidth to multiple bands, for instance continuous and random hopping bands, for better detection accuracies. It can be also noticed that the *oclass* detection accuracies are quite good even in the worst performing band 0 showing the robustness of the signal classification module of the encoder.

The anomaly detection accuracies of SAIFE are also compared against other algorithms on the SDR dataset as shown in Figure 9. On the high SNR regions OSVM and LODA gives a better or close enough performance as SAIFE. On the low SNR regions SAIFE performs better. Once again SAIFE adds much more interpretability to the whole anomaly detection process which is a huge advantage.

C. Anomaly Localization

Localizing anomalies in the wireless frequency spectrum is not common in any of the SoA algorithms. SAIFE presents a simple and robust way to localize the anomalous region from the input PSD data which is a significant contribution of this paper. In addition to detection of anomalies, the reconstruction error along with the semi-supervised features can be used to localize and understand the anomaly better as shown in Figure 12. Anomaly localization is achieved by plotting the absolute reconstruction error, that is $|\hat{\mathbf{x}}-\mathbf{x}|$. This method works well unless there is a drastic change in the estimated class label which can be noticed in the third row where an *scont* anomaly occurs in an *srshort* band. The model accurately detects it as an anomaly since there is a variation in the estimated class label, but shows the *srshort* signal as the anomalous region instead of *scont*. Figure 14 gives some sample plots of the estimated signal position and bandwidth. The current model is only trained for three semi-supervised features including the class label and these interpretable features can be used for analyzing the anomalies better.

D. Anomaly Detection in the Wild

To understand the performance on detecting real anomalies, the model is tested on the real-world Electrosense dataset. The model is trained on 7 days of data from one of the Electrosense sensors and tested on the next 500 hours for anomalies with a detection threshold of 3σ ($n = 3$). The number of detected anomalies, based on A_{score}, along with

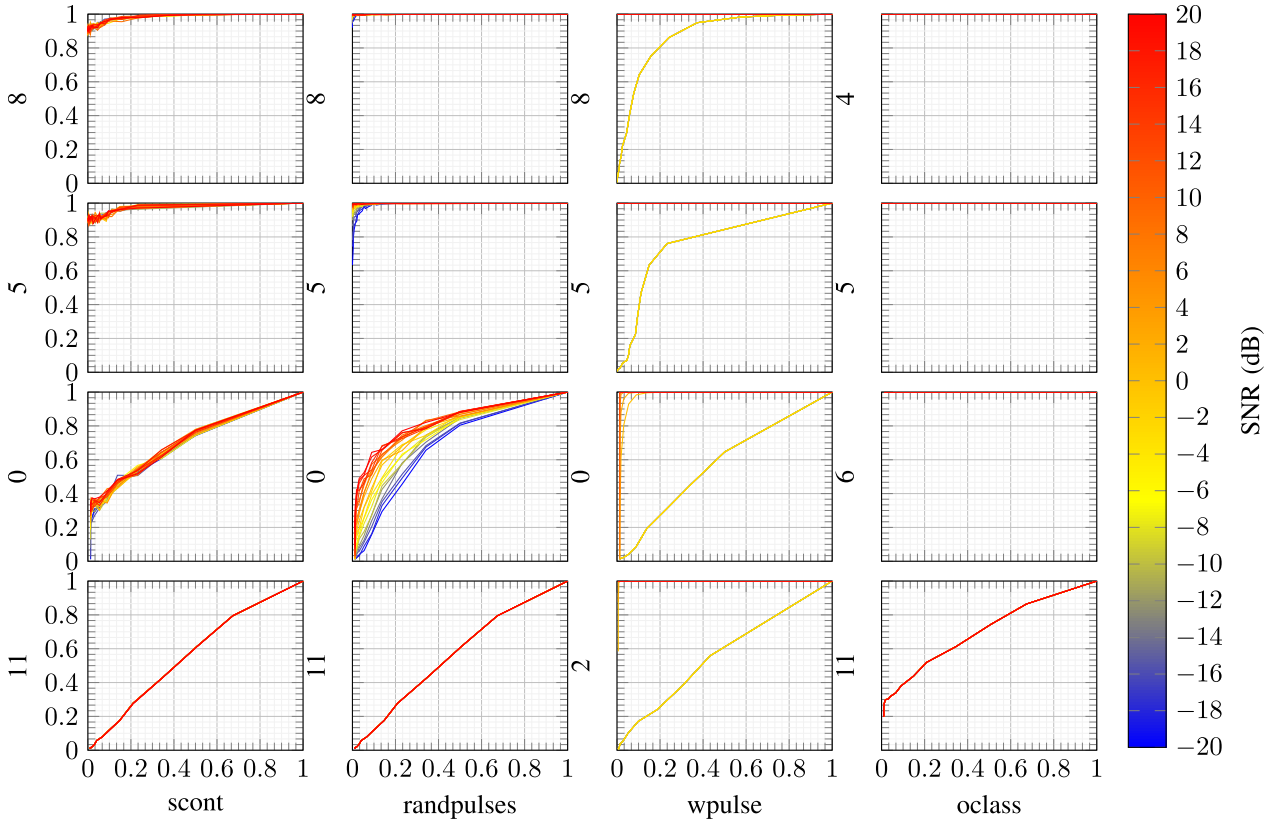

Fig. 11. ROC curves for the best two (top rows) and worst two (bottom rows) ROC AUC for −20dB SNR in SDR dataset. Synthetic bands with band labels are shown in rows for each synthetic anomaly (columns). On each plot false positive rate is represented on x-axis and true positive rate on y-axis. For *oclass* anomaly, anomaly vectors are randomly selected from other classes and specific SNR based ROC curves are not plotted.

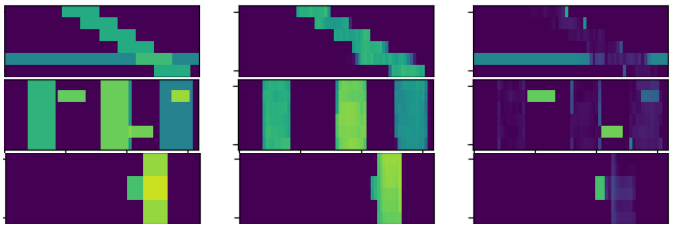

Fig. 12. Localized anomalies for three different synthetic anomalies. Original input signal, decoder reconstructed signal and the localized anomaly is shown in each row from left to right. First row: *wpulse* anomaly on *dethop* signal, second row: *randpulses* anomaly on *mult-cont* signal, third row: *scont* anomaly on *single-rshort* signal.

a few sample anomalies for 7 frequency bands are shown in Figure 13. The model detects unexpected missing transmissions (top-right and bottom-right), high power transmissions (bottom-left) and some out of band transmissions (top-left). It can be noticed that after 230 hours the 192-197 MHz bands started giving more anomalous detections. Visual inspection of the anomalous PSD patches in this band revealed transmission pattern variations. These detected variations could be either because of the transmitter behavior changes or from the position/antenna changes of the sensor.The model also provides the flexibility to add these anomalous detections to the training set, enabling incremental learning, if the user believes that the behavior is normal. The model retraining complexity is quite moderate as detailed in Section IV-A. Incorporating this user feedback and enabling automated retraining of models on these kind of anomalous behaviors will be addressed in future work.

VI. Signal Compression and Classification

To control the data transfer costs associated with the sensing, Electrosense sensors enable three pipelines with very low, medium and high data transfer costs namely: Feature, PSD and IQ pipeline. While the IQ pipeline allows to send raw data to the backend, which can be used to support a broad range of applications, the data transfer rate required is in the 30 Mbps to 100 Mbps range based on the sampling rate of the sensor. The PSD pipeline on the other hand brings down this rate to hundreds of Kbps. In this section we analyze the compression and classification capabilities of SAIFE to reduce the associated data transfer costs.

A. Traditional Spectrum Representation

In spite of the popularity of various lossy and lossless compression algorithms in image and video processing communities, there are only a few compression algorithms fine tuned for wireless spectrum data. In [22] the authors presented a compression algorithm based on Chebyshev polynomials. The authors in [23] presented a method to separate spectrum noise and other relevant signals specific to L-band satellite signals and then did separate compression to achieve better results

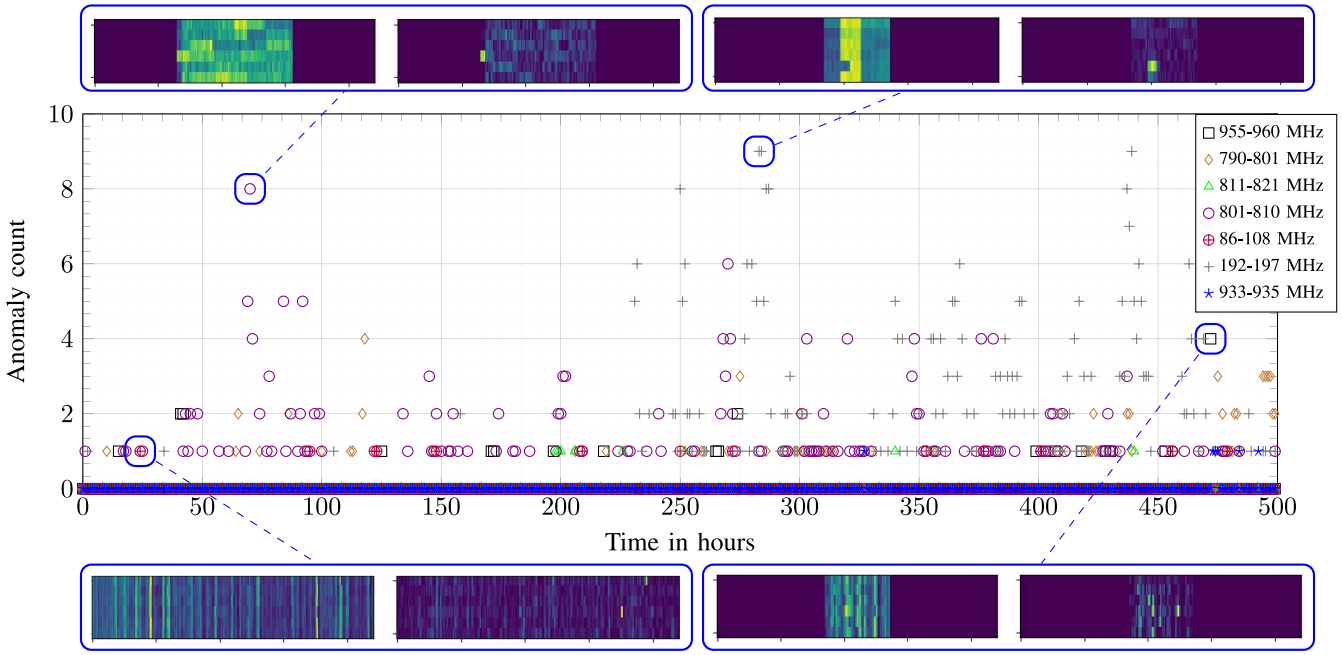

Fig. 13. Detected anomalies for a duration of 500 hours from one of the Electrosense sensor. Sample input data (left) and the localized anomaly (right) for some sample anomalies are also plotted for some frequency bands.

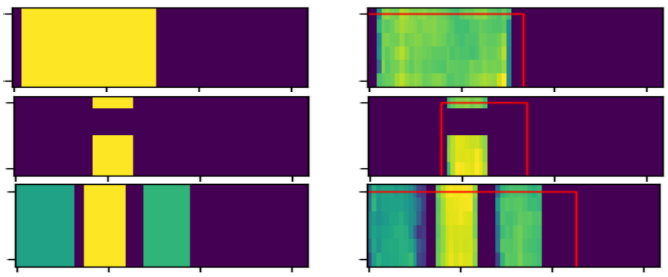

Fig. 14. Learned bandwidth and position features in a semi-supervised fashion. On each row the left one is the input signal and the right one is the reconstructed signal along with the estimated parameters.

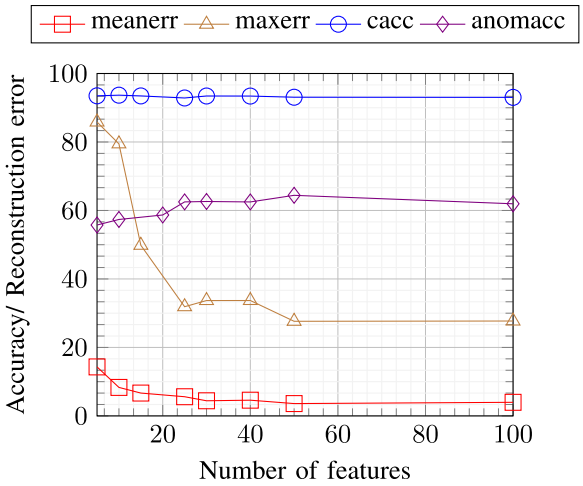

Fig. 15. Overall averaged anomaly detection accuracy (anomacc), wireless frequency band classification accuracy (cacc), mean absolute reconstruction error and maximum absolute reconstruction error of SAIFE on the synthetic dataset for a false alarm rate of 10%.

when compared to JPEG standards. The aforementioned methods are very specific and lack the compression flexibility when the input data is in multiple formats such as PSD or IQ.

B. Non-Linear Data Compression

Recently unsupervised deep learning models have shown great improvements in compressing input information. In [24] the authors have achieved 4X to 16X compression ratio on raw sampled IQ data using autoencoder models. In SAIFE, 20 compressed features are used for representing the input PSD frame. This helps to achieve a lossy compression of 19X, 60X and 120X on the Synthetic, Electrosense and SDR datasets respectively, which can considerably reduce the data transfer costs. Mean absolute reconstruction error of SAIFE with 20 features are summarized in Table IV. In addition to spectrum reconstruction these features can be used for anomaly detection and signal classification which makes it more attractive. The models can be easily adapted for different data inputs, for instance PSD data in time and frequency or IQ data supporting flexible compression architectures for different sensor data pipelines. The dimension of the compressed feature space along with the model complexity can be adapted to suit the reconstruction loss requirements. For instance, the number of features required to represent the time-frequency PSD patches of static wireless channels like commercial FM bands will be very less when compared to very random hopping channels.

An initial analysis is performed, on the synthetic dataset, to understand the trade-off between level of compression and anomaly detection performance by varying the number of continuous features, thereby the compression ratio of the model, which is presented in Figure 16. At very low (-20 dB) and high (20 dB) anomaly SNRs there are not much performance gains by increasing the number of features as

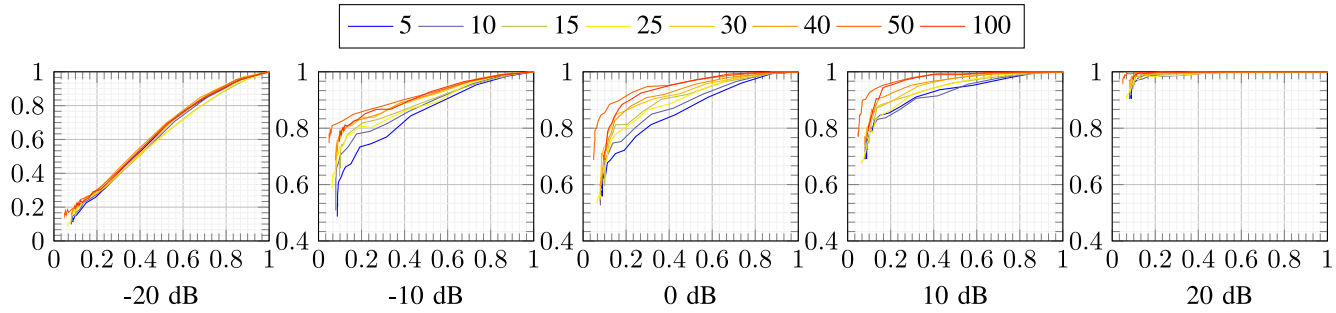

Fig. 16. ROC curves for *single-rshort* signal with *scont* anomalies at different SNRs with varying number of continuous features.

TABLE IV
BAND CLASSIFICATION ACCURACY AND RECONSTRUCTION ERRORS ON THE TEST DATA OF DIFFERENT DATASETS

Dataset	Classification accuracy (%)	Mean absolute reconstruction error
Synthetic dataset	92.86	7.67 (for 6x64 samples)
SDR dataset	100	84.12 (for 6x400 samples)
Electrosense dataset	100	101.16 (for 6x221 samples)

expected. Detecting signals at −20 dB SNR is very difficult even with a large number of features whereas at 20 dB SNR with a smaller set of features can easily detect anomalies due to large variations in the reconstruction loss. While at common SNR values (−10dB, 0 and 10dB) the anomaly detection performance increases with increasing number of features. We would like to emphasize that the number of features required to achieve reasonable detection performance will depend on the input data dimensions, the encoder and decoder capacity and the dataset complexity itself.

Figure 15 presents the variations in the overall anomaly detection accuracy along the maximum and mean reconstruction error and the wireless frequency band classification performance with increasing number of intermediate features. It can be noticed that the reconstruction error reduces steadily till one point (50 features) and then saturates afterwards as expected. We strongly believe that lower reconstruction loss can be achieved by using very deep and robust CNN models which are successfully used in high resolution image reconstruction problems. For these future models, even with lower reconstruction error we expect a similar saturation with increasing number of features. The lower anomaly detection accuracies (around 60%) is attributed to the averaging of accuracies over different bands, anomalies and SNRs. The wireless band classification accuracy is not that affected by varying number of features as one of the features is *class* which is trained in a semi-supervised fashion.

C. Wireless Signal Classification

In addition to anomaly detection ROC curves, wireless band classification accuracies on the test data of three datasets are summarized on Table IV. Since the real wireless bands use different parameters such as signal bandwidths, modulation type, and temporal occupancies at mostly high SNRs, the wireless band classification problem is not very tough as the classical modulation classification problem [2]. On the synthetic dataset

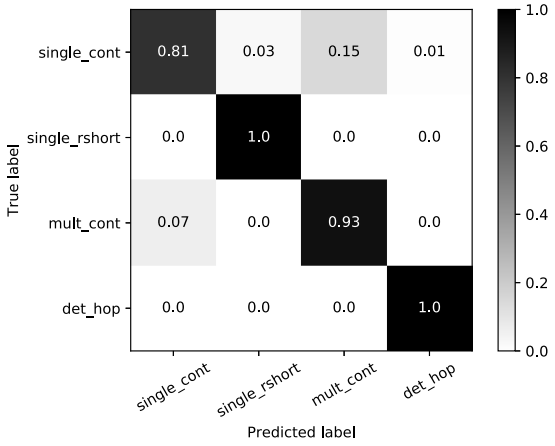

Fig. 17. Confusion matrix for the synthetic dataset.

the model confuses between *single-cont* and *mult-cont* signal resulting in a classification accuracy of 92.86%. The confusion matrix for the same is also shown in Figure 17. The model achieves excellent classification accuracy of 100% on the real SDR and Electrosense dataset. The high classification accuracy stress the fact that a categorical variable helps the encoding process which in-turn helps the decoder to generate fine variations which are specific to a particular class.

VII. CONCLUSION AND FUTURE WORK

Automated monitoring of wireless spectrum over frequency, time and space is still a difficult research problem. In this paper we have analyzed the use of an AAE model in wireless spectrum data anomaly detection, compression and signal classification. We have shown that the proposed model can achieve good anomaly detection and localization along with interpretable feature extraction. The model also can achieve a wireless band classification accuracy close to 100% by only using 20% labeled samples. Further, the performance of the proposed model is compared against various SoA anomaly detection algorithms in literature showing its robustness.

In future we would like to perform detailed comparisons of the proposed model with similar prediction based models and also evaluate the performance gains by using raw IQ samples. Even though we have validated the model performance on one of the Electrosense sensors, we would like to propose some concrete similarity scores that can be used to select closely

located or similar spectrum scanning sensors, to enable deployment of a single model across sensors. Further we would like to include user feedback in the entire anomaly detection loop and make the training process fully automated to fulfill the automated spectrum monitoring dream.

REFERENCES

[1] S. Rajendran *et al.*, "Electrosense: Open and big spectrum data," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 210–217, Jan. 2018.

[2] 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. Cogn. Commun. Netw.*, vol. 4, no. 3, pp. 433–445, Sep. 2018.

[3] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," *IEEE Wireless Commun.*, vol. 15, no. 4, pp. 34–40, Aug. 2008.

[4] M. Xie, S. Han, B. Tian, and S. Parvin, "Anomaly detection in wireless sensor networks: A survey," *J. Netw. Comput. Appl.*, vol. 34, no. 4, pp. 1302–1325, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804511000580

[5] A. Patcha and J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Comput. Netw.*, vol. 51, no. 12, pp. 3448–3470, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S138912860700062X

[6] S. Liu, Y. Chen, W. Trappe, and L. J. Greenstein, "ALDO: An anomaly detection framework for dynamic spectrum access networks," in *Proc. IEEE INFOCOM*, Apr. 2009, pp. 675–683.

[7] W. Honghao, J. Yunfeng, and W. Lei, "Spectrum anomalies autonomous detection in cognitive radio using hidden Markov models," in *Proc. IEEE Adv. Inf. Technol. Electron. Autom. Control Conf. (IAEAC)*, Dec. 2015, pp. 388–392.

[8] T. J. O'Shea, T. C. Clancy, and R. W. McGwier, "Recurrent neural radio anomaly detection," *arXiv e-prints*, Nov. 2016. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv161100301O

[9] N. Tandiya, A. Jauhar, V. Marojevic, and J. H. Reed, "Deep predictive coding neural network for RF anomaly detection in wireless networks," *arXiv e-prints*, Mar. 2018. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2018arXiv180306054T

[10] Q. Feng, Y. Zhang, C. Li, Z. Dou, and J. Wang, "Anomaly detection of spectrum in wireless communication via deep auto-encoders," *J. Supercomput.*, vol. 73, no. 7, pp. 3161–3178, 2017.

[11] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv e-prints*, Dec. 2013. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6114K

[12] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv e-prints*, Nov. 2015. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2015arXiv151105644M

[13] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[14] X. Chen *et al.*, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," *arXiv e-prints*, Jun. 2016. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160603657C

[15] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," *arXiv e-prints*, Mar. 2016. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160304467A

[16] *GNU Radio Website*. Accessed: May 5, 2018. [Online]. Available: http://www.gnuradio.org

[17] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *arXiv e-prints*, Sep. 2016. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2016arXiv160907061H

[18] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv e-prints*, Dec. 2014. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1412.6980K

[19] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation-based anomaly detection," *ACM Trans. Knowl. Disc. Data*, vol. 6, no. 1, p. 3, 2012.

[20] T. Pevnỳ, "Loda: Lightweight on-line detector of anomalies," *Mach. Learn.*, vol. 102, no. 2, pp. 275–304, 2016.

[21] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.

[22] S. E. Hawkins, III, E. H. Darlington, A. F. Cheng, and J. R. Hayes, "A new compression algorithm for spectral and time-series data," *Acta Astronautica*, vol. 52, nos. 2–6, pp. 487–492, 2003.

[23] Y. Li, Z. Gao, L. Huang, Z. Tang, and X. Du, "A wideband spectrum data segment compression algorithm in cognitive radio networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[24] T. J. O'Shea, J. Corgan, and T. C. Clancy, "Unsupervised representation learning of structured radio communication signals," in *Proc. IEEE 1st Int. Workshop Sens. Process. Learn. Intell. Mach. (SPLINE)*, 2016, pp. 1–5.

Sreeraj Rajendran received the master's degree in communication and signal processing from the Indian Institute of Technology, Bombay, in 2013. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering, KU Leuven, Belgium. He was a Senior Design Engineer with the Baseband Team of Cadence and an ASIC Verification Engineer with Wipro Technologies. His main research interests include machine learning algorithms for wireless and low power wireless sensor networks.

Wannes Meert received the Master of Electrotechnical Engineering degree in microelectronics, the Master of Artificial Intelligence degree, and the Ph.D. degree in computer science from KU Leuven in 2005, 2006, and 2011, respectively, where he is currently a Research Manager with DTAI Research Group. His work is focused on applying machine learning, artificial intelligence, and anomaly detection technology to industrial application domains.

Vincent Lenders (M'05) received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering and information technology from ETH Zurich, Switzerland, in 2001 and 2006, respectively. He is the Head of C4I Networks Group and Cyber-Defence Campus with armasuisse. He is also the Co-Founder and the Chairman of the executive boards of the OpenSky Network and Electrosense Associations. He was also a Post-Doctoral Research Faculty with Princeton University, USA. He has authored over 100 publications that appeared in peer-reviewed international conferences and journals and invented two patents. He was a recipient of the Best Paper Awards at IEEE WONS 2012, DFRWS EU 2015, ACM CPSS 2015, and DASC 2015, and the Security Award in 2011 from the Swiss Federal Department of Defense. He is a member of ACM, and the expert Jury of the Swiss Economic Forum. He holds various security professional and auditor certifications, including CISA, CISM, CRISC (ISACA), and CISSP (ISC2).

Sofie Pollin (S'02–M'06–SM'13) received the Ph.D. degree (Hons.) from KU Leuven in 2006. From 2006 to 2008, she continued her research on wireless communication, energy-efficient networks, cross-layer design, coexistence, and cognitive radio with the University of California at Berkeley. In 2008, she returned to imec to become a Principal Scientist with the Green Radio Team. Since 2012, she has been a Tenure Track Assistant Professor with the Electrical Engineering Department, KU Leuven. Her research centers around networked systems that require networks that are ever more dense, heterogeneous, battery powered, and spectrum constrained. She is a fellow of BAEF and Marie Curie.