Citation: Gusain, Kuanl, Hassan, Zoheb, Couto, David, Malek, Mai Abdel, Shah, Vijay K et al. 2024. "Automated and Blind Detection of Low Probability of Intercept RF Anomaly Signals."

As Published: https://doi.org/10.1145/3636534.3698243

Publisher: ACM|The 30th Annual International Conference on Mobile Computing and Networking

Persistent URL: https://hdl.handle.net/1721.1/158078

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Massachusetts Institute of Technology

Automated and Blind Detection of Low Probability of Intercept RF Anomaly Signals

Kunal Gusain
Microsoft
Redmond, WA, USA
kunalg@vt.edu

Zoheb Hassan
Laval University
Quebec City, QC, Canada
MDHAS1@ulaval.ca

David J. Couto
BAE Systems Inc.
Merrimack, NH, USA
david.couto@baesystems.com

Mai A. Abdel-Malek
University of Arizona
Tucson, USA
mmalek@arizona.edu

Vijay K. Shah
North Carolina State University
Raleigh, NC, USA
vijay.shah@ncsu.edu

Lizhong Zheng
Massachusetts Institute of Technology
Cambridge, MA, USA
lizhong@mit.edu

Jeffrey H. Reed
Virginia Tech
Blacksburg, VA, USA
reedjh@vt.edu

Abstract

Automated spectrum monitoring necessitates the accurate detection of low probability of intercept (LPI) radio frequency (RF) anomaly signals to identify unwanted interference in wireless networks. However, detecting these unforeseen low-power RF signals is fundamentally challenging due to the scarcity of labeled RF anomaly data. In this paper, we introduce WANDA (Wireless ANomaly Detection Algorithm), an automated framework designed to detect LPI RF anomaly signals in low signal-to-interference ratio (SIR) environments without relying on labeled data. WANDA operates through a two-step process: (i) Information extraction, where a convolutional neural network (CNN) utilizing soft Hirschfeld-Gebelein-Rényi correlation (HGR) as the loss function extracts informative features from RF spectrograms; and (ii) Anomaly detection, where the extracted features are applied to a one-class support vector machine (SVM) classifier to infer RF anomalies. To validate the effectiveness of WANDA, we present a case study focused on detecting unknown Bluetooth signals within the WiFi spectrum using a practical dataset. Experimental results demonstrate that WANDA outperforms other methods in detecting anomaly signals across a range of SIR values (-10 dB to 20 dB).

1 Introduction

Low probability of intercept (LPI) signals represent signals that are designed to conceal sensitive information by transmitting noise-like waveforms using direct-sequence spread spectrum (DSSS) and frequency hopping, rendering the transmitted signals practically undetectable or unrecognizable to external wireless networks [1]. Accurate detection LPI RF anomaly signals is required to obtain critical and real-time situational awareness in tactical wireless networks, public safety networks, and security-sensitive cyber-physical systems. Detection of RF anomaly signals relies on extracting specific signal features from the input observations (e.g., raw in-phase and quadrature signals and spectrograms) such that these features effectively distinguish known and unknown (anomaly) signals. Conventional algorithms rely on manually crafted pattern extraction techniques [2–4] for anomaly signal detection. *The key challenge in detecting LPI signals is that the LPI RF signals are often obscured by noise, causing the required signal features to be lost*. The conventional cyclo-stationary signal processing (CSP) signal processing schemes, despite the capability of extracting useful signal features in low signal-to-noise ratio (SNR) regimes, require high computational complexity, long signal collection duration, and signal-specific expert domain knowledge for LPI signal detection [5].

Modern approaches surpass traditional algorithms by leveraging deep learning (DL) to extract meaningful features from spectrograms. DL algorithms can compare features extracted from an *unlabeled* spectrogram with those of a known signal to identify potential anomalies. State-of-the-art methods like Autoencoders (AE) and recurrent neural networks (RNN) are applied to this task. AEs, for example, often fail to accurately reconstruct spectrograms from features at the bottleneck layer when anomalies are present, resulting in large reconstruction errors that imply the presence of unknown elements in the signal space [6, 7]. Similarly, RNNs predict future observations based on previous ones, with significant prediction errors typically indicating deviations from known patterns, thereby identifying anomalies [8, 9]. For instance, a deep video prediction network, PredNet, was employed in [10] to predict spectrograms and use prediction

error as a metric for anomaly detection. Convolutional neural networks (CNNs) are also proven effective in detecting RF anomalies. In [11], a 1D CNN was used to detect LPI RF interference from I/Q samples, while [12] combined CSP features and I/Q samples at the CNN input to enhance DSSS signal detection in the LTE band.

However, state-of-the-art DL approaches for detecting LPI RF anomaly signals face several challenges. First, the reconstruction and prediction of high-dimensional spectrograms, as required in AE and RNN approaches, are challenging tasks, especially in highly dynamic and noisy signal propagation environments. Second, the CNN approaches require supervised training with manually labeled RF anomaly signals. However, obtaining labeled spectrograms for every possible anomaly within a given band is highly impractical, if not impossible. Lastly, these existing DL methods may not always extract useful low-dimensional features (that differentiate the known and anomaly signals) from high-dimensional spectrograms by filtering out noise and other unwanted information. This limitation reduces their efficiency in detecting unknown LPI RF anomalies in the spectrum. Consequently, there is a need for novel unsupervised DL approaches capable of detecting LPI RF anomaly signals without relying on preexisting labels.

Contributions: In this work, we introduce the WANDA framework, a novel solution aimed at addressing the challenge of automated and blind LPI RF anomaly signal detection. Our goal is to significantly improve the network's capability to detect subtle and unknown RF signals within the spectrum. WANDA (i) extracts informative features required for anomaly detection by filtering out noise and other unwanted information from spectrograms; (ii) facilitates an end-to-end unsupervised training approach without requiring any labeled RF anomaly signals during training; and (iii) does not require any prior knowledge of SNRs or the decoding/demodulating of received signals for anomaly detection. Combining these attributes, our proposed WANDA framework proves highly effective in detecting unseen and unlabeled LPI RF anomaly signals. The specific contributions of this work are summarized as follows.

• A DL-enabled RF anomaly signal detection system, called by WANDA (**W**ireless **AN**omaly **D**etection **A**lgorithm) is proposed. WANDA monitors the RF spectrum, extracts information from the observations, and detects the presence of anomaly signal(s) as shown in Fig. 1. To conduct these steps, efficient feature extraction and anomaly detection engines are designed for WANDA.

• A novel CNN is designed for the feature extraction engine of WANDA. More precisely, the proposed CNN employs maximal soft Hirschfeld-Gebelein-Rényi (HGR) correlation or H-score (discussed in Section 3) as the loss function.

• A one-class support vector machine (OC-SVM) classifier is employed in the anomaly detection engine of WANDA. The OC-SVM utilizes features extracted by the feature extraction engine to classify spectrograms as either clean or anomalous. The OC-SVM classifier is trained by considering the features of known signals, which are obtained by processing the known signal's spectrograms to the feature extraction engine, as the *positive class* features.

• To assess the performance of WANDA, a simulation case study is designed to accurately detect the presence of unknown low-power RF Bluetooth signals in the WiFi spectrum by leveraging a publicly available dataset. Experimental results confirm that WANDA
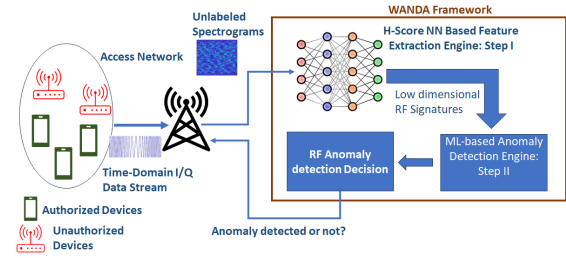

Figure 1: Overview of WANDA framework.

achieves a higher true probability of detection and better RF anomaly detection accuracy compared to several benchmark schemes.

2 Overview of WANDA

As shown in Fig. 1, the network controller periodically collects in-phase/quadrature (I/Q) samples from the RF environment (with both authorized and unauthorized devices). These samples are then converted into spectrograms and fed into the WANDA framework. WANDA performs real-time analysis to determine the presence of an RF anomaly signal and communicates its decision to the network controller. This work primarily focuses on accurately detecting low-power and unknown RF anomaly signals that are present within the spectrum, and the mitigation strategies for handling such unknown interference are out of the scope of this work.

Now, let's delve into the intricacies of the proposed WANDA framework. WANDA consists of two sequential steps, namely, the information extraction and anomaly detection steps. The anomaly detection from spectrogram is analogous to a hidden Markov model [13, Ch. 5]. In particular, the observations for our considered problem is a sequence of high-dimensional spectrograms, denoted by $\{\mathcal{Z}_n, n = 0, 1, \cdots\}$. Such spectrograms temporally vary due to the (hidden) activity factor(s) of anomalous source, denoted by $\{\bar{z}_n \in \{\text{ON}, \text{OFF}\}, n = 0, 1, \cdots\}$. The goal is to learn the mapping, $\mathcal{Z}_n \to \mathbf{z}_n, n = 0, 1, \cdots$, without any further knowledge of the statistical model (e.g., state transition probability matrix of $\mathbf{z}_n$ and the conditional probability of $\mathcal{Z}_n$ given $\mathbf{z}_n$). Since spectrogram is usually a high-dimensional 2D image, it contains both information-of-interest and information-of-not-interest. As a result, learning of $\{\mathbf{z}_n\}$ would be more efficient from $\{\hat{Z}_n\}$ rather than from $\{\mathcal{Z}_n\}$, where $\{\hat{Z}_n\}$ is a set of carefully extracted features by removing background noise and other information-of-not-interest. Accordingly, we consider learning the following mapping $\mathcal{Z}_n \to \hat{Z}_n \to \mathbf{z}_n, n = 0, 1, \cdots$. WANDA performs such a mapping using the following two steps.

• **Step I** (learning $\mathcal{Z}_n \to \hat{Z}_n$): Here, WANDA extracts a set of low-dimensional features from high-dimensional spectrograms of the input signal as such these features contain information about the spectrogram and distinguish spectrograms of known and anomaly signals. A custom CNN with unique loss function, described in Section 3, is employed to conduct the feature extraction process.

• **Step II** (learning $\hat{Z}_n \to \mathbf{z}_n$): Here, an ML-based anomaly detection approach is employed to learn the mapping of RF anomaly signal's state in the spectrum from on the extracted features.

These two steps are sequentially trained offline, and both of them are entirely unsupervised, i.e., they do not require any manually labeled anomaly signals and features. The detailed description of Steps I and II of WANDA is explained in Sections 3 and 4, respectively.

3 WANDA Framework Design: Feature Extraction Engine (Step I)

The feature extraction engine of WANDA is a customized CNN (dubbed as H-score CNN henceforward) trained to extract RF signatures by removing noise from the spectrogram. Such a CNN is obtained by modifying its loss function to the soft-HGR maximal correlation function. Section 3. A presents Algorithm 1 to perform soft-HGR maximal correlation by leveraging an unsupervised DL method. Section 3. B explains our proposed CNN-based implementation of Algorithm 1 (i.e., the implementation of H-score CNN).

3.A Feature Extraction Algorithm Development.

Soft-HGR Correlation Approach. Let $X \in \mathcal{U}$ and $Y \in \mathcal{U}$ be a pair of spectrograms of the same signal at different time instants, with $\mathcal{U}$ being the universal set of spectrograms. Thus, X and Y have certain statistical correlations. The HGR maximal correlation between X and Y aims to find the non-linear transformations $\mathbf{f}^*(\cdot) = [f_1^*, f_2^*, \cdots, f_K^*]$ and $\mathbf{g}^*(\cdot) = [g_1^*, g_2^*, \cdots, g_K^*]$, known as the HGR maximal correlation functions, by solving the following optimization problem [14].

$$\mathbf{f}^*(\cdot), \mathbf{g}^*(\cdot) = \arg\max_{\mathbf{f},\mathbf{g}} \mathbb{E}\left[\mathbf{f}^T(X)\mathbf{g}(Y)\right]$$
$$\text{s.t.} \quad \mathbb{E}\left[f_i(x)\right] = \mathbb{E}\left[g_i(x)\right] = 0, i = 1, 2, \cdots, K-1$$
$$\mathbb{E}\left[f_i^*(x)f_j^*(x)\right] = \mathbf{1}_{i=j}, i, j \in \{1, 2, \cdots, K-1\}$$
$$\mathbb{E}\left[g_i^*(x)g_j^*(x)\right] = \mathbf{1}_{i=j}, i, j \in \{1, 2, \cdots, K-1\}$$

(1)

where $\mathbf{f}(x) = [f_1(x), \cdots, f_K(x)]^T$ and $\mathbf{g}(x) = [g_1(x), \cdots g_K(x)]^T$ are known as the feature vectors and $K = |\mathcal{U}|$. The motivation for considering HGR maximal correlation problem (i.e., eq. (1)) is explained as follows. The associated maximal correlation between the k-th feature vectors is obtained as $\sigma_k = \mathbb{E}\left[f_k^*(X)g_k^*(Y)\right], k \in \{1, 2, \cdots, K-1\}$, and the maximal HGR correlation between X and Y is obtained as $H(X, Y) = \sum_{k=1}^{K} \sigma_k$. Moreover, such HGR correlation is related to the mutual information between X and Y as $I(X, Y) \approx \frac{1}{2}\sum_{k=1}^{K} \sigma_k^2$, where $I(X, Y)$ is the mutual information between X and Y [14]. Essentially, the feature vectors obtained from the HGR maximal correlation problem capture the maximum mutual information between two spectrograms. *Said differently, these features encapsulate certain information about the spectrograms. Note that the spectrograms of an RF signal exhibit unique patterns that depend on the signal type. HGR correlation can extract features capturing these distinctive patterns without any prior labels.*

However, HGR maximal correlation approach requires high computational complexity for feature extraction from high-dimensional observations [15]. To overcome such an impediment, the following optimization problem [16, eq. (17)] can be considered to alternatively determine the optimal feature functions $\mathbf{f}^*(\cdot)$ and $\mathbf{g}^*(\cdot)$.

$$\mathbf{f}^*(\cdot), \mathbf{g}^*(\cdot) = \arg\max_{\mathbf{f},\mathbf{g}} \mathbb{E}\left[\mathbf{f}^T(X)\mathbf{g}(Y)\right] - \frac{1}{2}\text{tr}\left(\text{cov}\left(\mathbf{f}(X)\right)\text{cov}\left(\mathbf{g}(Y)\right)\right)$$
$$\text{s.t.} \quad \mathbb{E}[\mathbf{f}(X)] = \mathbb{E}[\mathbf{g}(Y)] = 0.$$

(2)

Algorithm 1 Proposed Information Extraction Algorithm

Input: Paired spectrograms in an m-size mini-batch: $(\boldsymbol{x}^{(1)}, \boldsymbol{y}^{(1)}), (\boldsymbol{x}^{(2)}, \boldsymbol{y}^{(2)}), \cdots, (\boldsymbol{x}^{(m)}, \boldsymbol{y}^{(m)})$.
Initialize: Neural Network Parameters θ.
repeat
    Compute feature functions, $\mathbf{f}_\theta(\boldsymbol{x}^{(i)})$ and $\mathbf{g}_\theta(\boldsymbol{y}^{(i)})$, $\forall i = 1, 2, \cdots, m$.
    Compute normalized feature functions:

$$\mathbf{f}_\theta(\boldsymbol{x}^{(i)}) \leftarrow \mathbf{f}_\theta(\boldsymbol{x}^{(i)}) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{f}_\theta(\boldsymbol{x}^{(i)}), \forall i = 1, 2, \cdots, m$$

$$\mathbf{g}_\theta(\boldsymbol{y}^{(i)}) \leftarrow \mathbf{g}_\theta(\boldsymbol{y}^{(i)}) - \frac{1}{m}\sum_{i=1}^{m}\mathbf{g}_\theta(\boldsymbol{y}^{(i)}), \forall i = 1, 2, \cdots, m$$

    Compute the sample covariance:

$$\text{cov}(\mathbf{f}) \leftarrow \frac{1}{m}\sum_{i=1}^{m}\mathbf{f}_\theta(\boldsymbol{x}^{(i)})\mathbf{f}_\theta(\boldsymbol{x}^{(i)})^T$$

$$\text{cov}(\mathbf{g}) \leftarrow \frac{1}{m}\sum_{i=1}^{m}\mathbf{g}_\theta(\boldsymbol{y}^{(i)})\mathbf{g}_\theta(\boldsymbol{y}^{(i)})^T$$

    Compute the H-score:

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{f}_\theta(\boldsymbol{x}^{(i)})\mathbf{g}_\theta(\boldsymbol{y}^{(i)}) - \frac{1}{2}\text{tr}\left(\text{cov}(\mathbf{f})\text{cov}(\mathbf{g})\right)$$

    Update the neural network parameters, θ, by considering H-score as the loss function and applying the SGD and backpropagation techniques.
until Convergence or maximum number of iterations are reached
Output: Feature functions $\mathbf{f}_{\theta*}(\cdot)$ and $\mathbf{g}_{\theta*}(\cdot)$

Eq. (2) is known as the soft-HGR maximal correlation problem, and its objective function is defined as the **H-score**. Similar to maximizing HGR correlation, maximizing H-score also allows us to obtain the feature vectors having maximum mutual information content between X and Y, with the reduced computational complexity [16]. Hence, *we exploit the soft-HGR maximal correlation approach to extract the RF signature from high-dimensional spectrograms.*

Unsupervised DL algorithm to solve (2). Eq. (2) is an infinite dimensional optimization problem over the function space, and as a result, it is highly challenging to solve this problem optimally. To this end, we apply unsupervised DL to solve (2). More specifically, we introduce a DNN, parameterized by θ, to approximate the feature functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$. As a result, eq. (2) is equivalently written as

$$\theta^* = \arg\max_{\theta} \mathbb{E}\left[\mathbf{f}_\theta^T(X)\mathbf{g}_\theta(Y)\right] - \frac{1}{2}\text{tr}\left(\text{cov}\left(\mathbf{f}_\theta(X)\right)\text{cov}\left(\mathbf{g}_\theta(Y)\right)\right)$$
$$\text{s.t.} \quad \mathbb{E}[\mathbf{f}_\theta(X)] = \mathbb{E}[\mathbf{g}_\theta(Y)] = 0$$

(3)

where $\mathbf{f}_\theta(\cdot)$ and $\mathbf{g}_\theta(\cdot)$ denote the parameterized feature functions. The DNN takes the negative of the H-score as the loss function. We apply stochastic gradient descent (SGD) and backpropagation techniques to update the parameters of H-score NN. The overall algorithm for determining optimal feature functions by solving (3) is summarized as Algorithm 1. The computational complexity of Algorithm 1 is obtained as $O(mK^2)$, where m and K are the total number of sample spectrograms and extracted features, respectively.

3.B Implementation of Feature Extraction Algorithm.
A multi-layer CNN is used to extract low-dimensional informative features from the spectrograms. At first, the stream of complex I/Q samples (i.e., time-domain signals) is decomposed into multiple frames, and spectrograms of each frame of I/Q samples are determined. The subsequent time-indexed spectrograms are passed in pair to the CNN and its parameters are updated using Algorithm 1. For CNN, we utilize two convolution layers with a ReLu activation function and a

max pool layer. We also use a fully connected layer with a sigmoid activation function before calculating the loss function. *In contrast to conventional approach, our proposed approach non-intuitively looks at the channels (i.e., filters) of the second-to-last layer of the trained CNN to extract the required features. Such an approach is motivated by the fact that after getting appropriately trained using Algorithm 1, the CNN learns to remove the non-required background noise from the spectrogram and retains only the portions of the spectrogram that contain signal specific information.* The configuration of the adopted CNN model for the simulation experiment is summarized in the table 1. We emphasize that the architecture and complexity of the required CNN model can vary according to the types of known signals. The reported CNN configuration is selected after several trials and error experiments.

Number of input samples, n	$3 \times 128 \times 128$
Size of minibatch	256
Optimizer	Stochastic gradient descent
Learning rate	e^{-5}
Epochs	200
Input shape	$3 \times 64 \times 64$
Kernel size	9×9
CNN output channels	8
Shape of the channel	16×16

Table 1: Configuration of CNN for Anomaly Detection

4 WANDA Framework Design: Anomaly Detection Engine (Step II)

Anomaly detection engine of WANDA performs mapping from the extracted features to the state of RF anomaly signal (e.g., ON/OFF or YES/NO). The overall procedure of training anomaly detection engine is described as follows. In particular, we first pass the spectrograms of known signals through the trained H-score CNN and obtain features from the CNN channels. Since the extracted features are essentially 2D matrices, we flatten them to vectors and train a one-class support vector machine (OC-SVM) classifier using such flattened feature vectors as the *positive class* features. The classification function used by the OC-SVM classifier for any given feature generated by the trained H-score CNN, denoted by $\mathbf{f}$, is expressed as $\Phi(\mathbf{f}) = \sum_{n=1}^{N} \alpha_n \mathcal{K}(\mathbf{f}, \mathbf{f}_n) - \rho$. Here, α_n and ρ are the classification parameters that are learned during offline training and $\mathcal{K}(\cdot)$ is the Kernel function. Different types of Kernel functions, such as linear, polynomial, and Gaussian radial basis functions, are available in the standard APIs, and it is a hyper-parameter of the OC-SVM classifier. The classification function is also known as the SVM score. During online inference, a trained OC-SVM classifier computes the SVM score for the feature vector(s) obtained by H-score CNN (i.e., Step I) and determines the presence and absence of anomaly signal if the SVM score is negative and positive, respectively. In order to obtain suitable anomaly detection accuracy, the hyper-parameters of the OC-SVM classifier need to be carefully optimized. To this end, we leverage the Optuna [17] framework, a state-of-the-art optimizer, to fine-tune the hyper-parameters of the OC-SVM classifier. In our simulation, we compare the anomaly detection performance of all the models (e.g., OC-SVM, ISOF, and SVDD) using their best hyper-parameters obtained by the Optuna framework.

5 Data Set Generation

Dataset Information. We use the CRAWDAD data set [18] to validate the effectiveness of WANDA in detecting the unknown

RF anomaly signals. The description of data set preparation is as follows. We first construct and reshape the dataset that emulates a continuous flow of raw signals. For simulation, we consider the CRAWDAD data set consisting of 225,000 signals, each consisting of 128 I/Q sample points [18]. More precisely, this data consists of 15 signal classes, with classes indexed by $1 - 10$ representing 1 MHz bandwidth IEEE 802.15.1 standard Bluetooth signals (with center frequencies from 2.422 GHz to 2.431 GHz), classes indexed by $11 - 13$ representing 20 MHz bandwidth IEEE 802.11 standard Wifi signals (with center frequencies 2.422 GHz, 2.427 GHz, and 2.432 GHz), and classes indexed by $14 - 15$ representing 2 MHz bandwidth IEEE 802.15.4 standard Zigbee signals (with center frequencies 2.425 GHz and 2.430 GHz). As mentioned earlier, we consider WiFi as our known signal and Bluetooth as the unknown and intermittently ON/OFF anomaly signal. Thus, we investigate the performance of our proposed WANDA framework to detect the presence of Bluetooth signals in the WiFi spectrum. To this end, we create training and testing data sets, where the training data set consists of only WiFi signals and the testing data set contains WiFi signals merged with randomly ON/OFF Bluetooth signals. For both training and testing data set generation, we consider four different SNR values, namely -10 dB, 0 dB, 10 dB, and 20 dB, for the WiFi signals while considering that the Bluetooth signal's SNR is 0 dB. In our simulation, we present the performance in terms of signal-to-interference (SIR) level in the dB unit, which is computed as $\mathbf{SIR}(dB) = \mathbf{SNR}_{WiFi}(dB) - \mathbf{SNR}_{BL}(dB)$ where $\mathrm{SNR}_{WiFi}(dB)$ and $\mathrm{SNR}_{BL}(dB)$ represent the SNR of WiFi and Bluetooth signals in the dB unit, respectively.

Spectrogram Generation. After creating the I/Q signal segments, we generate spectrograms from the created signals. We use the *pspectrum* function from MATLAB with the following set of parameters listed in Table 2. A total of 20k time-indexed spectrogram samples are generated for each signal-to-interference (SIR) value. The generated spectrograms are sequentially saved in an image format for all the time windows. The spectrograms generated at the i-th and $(i+N)$-th time windows are applied to the H-score CNN simultaneously, $\forall i$. This enables training the H-score CNN to extract features while exploiting statistical dependency between stream of spectrograms.

Parameter	Value
Minimum Threshold	-80 dB
Sampling Frequency	10^7 Hz
Overlap Percentage	99%
Leakage Ratio	1
Frequency Resolution	3×10^5 Hz

Table 2: MATLAB *pspectrum* function parameters

6 WANDA Experimental Evaluation

As mentioned in Section 5, a case study is designed to evaluate WANDA's performance by focusing on the task of detecting unknown or anomalous Bluetooth signals within the spectrum of WiFi signals, where WiFi serves as our known signal. We utilize the **AUC** (area under curve) and **ROC** (receiver operating characteristic curve) as our two key evaluation metrics. AUC provides valuable insights into how well our trained models perform at different thresholds by measuring the True Positive Rate (TPR) and False Positive Rate (FPR), defined as TPR $= \frac{TP}{TP+FN}$ and FPR $= \frac{FP}{TN+FP}$, respectively. Here, TP, TN, FP, and FN imply the total numbers of the
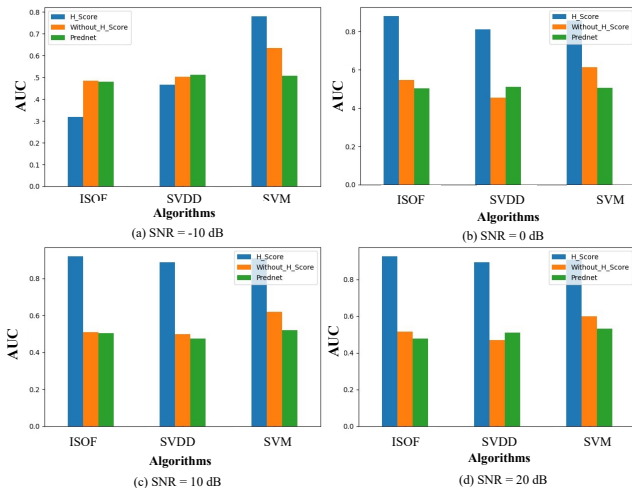
Figure 2: AUC performance of the proposed WANDA (denoted by H-score in the figures) and other benchmark schemes.

true predictions of the non-anomalous states, true predictions of the anomaly states, false predictions of the non-anomalous states, and false predictions of the anomaly states, respectively. A higher AUC value indicates an improved model's performance in distinguishing non-anomalous and anomalous samples. Meanwhile, the ROC curve visually represents the relationship between TPR and FPR by plotting them over X-axis and Y-axis, respectively, as different thresholds are considered. We consider the following two benchmark schemes for performance comparison.

Conventional Scheme (Without H-score CNN): This scheme infers the presence of the RF anomaly signal directly from the spectrogram using conventional anomaly detection approaches, namely, **OC-SVM**, **ISOF** (Isolation Forest), and **SVDD** (Support Vector Data Description), for detecting RF anomaly signals from the spectrogram. For training of all these approaches, we first convert the spectrogram of the WiFi signals into a 2D pixel matrix. Then, we flatten this matrix to a vector, reduce its dimension by applying principle component analysis, and use it as a feature input to the anomaly detector (i.e., OC-SVM, ISOF, or SVDD).

PredNet: In this scheme, we consider the prediction-based unsupervised RF anomaly detection scheme proposed in [10]. The RF anomaly signal detection is conducted in two steps. (I) In the **first** step, a trained deep video encoding network, called Prednet is utilized to predict the next spectrogram of the signal based on the observations of a certain number of previous spectrograms and a prediction error vector is computed. (II) In the **second** step, the computed prediction error vector is applied to an anomaly detection block in order to determine whether the spectrogram contains any RF anomaly signals or not. Similar to WANDA, both spectrogram prediction and anomaly detection stages are entirely unsupervised and trained using only WiFi spectrogram. In particular, for training the PredNet, a minimum mean square error criterion is considered to minimize the prediction error between the predicted and actual spectrogram images. Meanwhile, for training the anomaly detection block based on the prediction error vector, conventional RF anomaly detection schemes, such as OC-SVM, ISOF, and SVDD, are applied.
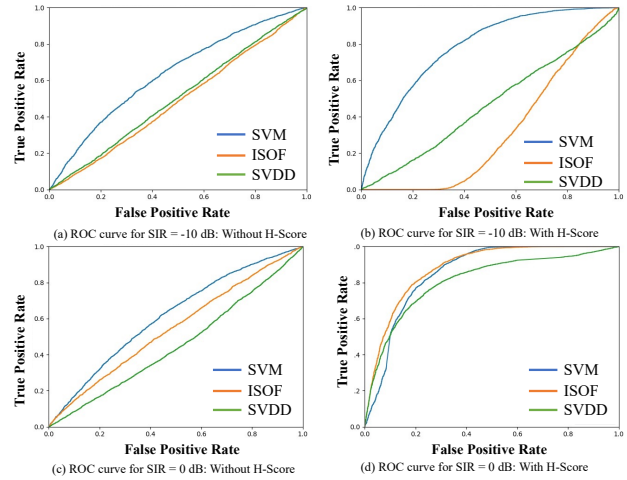

Figure 3: ROC Curves of Comparison for SIRs -10 and 0 dB.

6.A   Comparison With Benchmarks. Figs. 2(a)-2(d) compare the AUC score of proposed WANDA, PredNet, and conventional anomaly detection schemes. More specifically, in these figures, the legend "H-Score" and X-axis labels "SVM", ISOF" and "SVDD" imply the proposed WANDA framework, where the anomaly detection engine employs the OC-SVM, ISOF, and SVDD algorithms, respectively. Meanwhile, the legend "Without H-Score" and X-axis labels "SVM", ISOF" and "SVDD" imply the conventional anomaly detection schemes employing the OC-SVM, ISOF, and SVDD algorithms, respectively. Finally, the legend "PredNet" and X-axis labels "SVM", ISOF" and "SVDD" imply the PredNet scheme [10], where the anomaly detection step of PredNet employ the OC-SVM, ISOF, and SVDD algorithms, respectively.

Figs. 2(a)-2(d) show that Prednet achieves a poor AUC score in the range of 0.5 to 0.6 for all anomaly detection schemes, and this score does not significantly vary with SIRs. Recall, in the Pred-Net scheme the prediction error vector for the entire spectrogram is used for both training and testing anomaly detection blocks (i.e., OC-SVM, ISOF, and SVDD). However, in practice, RF anomaly signals are located only within a small portion of the spectrogram, and the remaining large portion of the spectrogram contains only background noise that is not impacted by the RF anomaly. As a result, the prediction error vectors for both known (i.e., Wifi) and anomaly impaired signals (e.g., Wifi with Bluetooth) looks almost same to the anomaly detection algorithms (e.g., OC-SVM, ISOF, and SVDD). Evidently, these algorithms cannot effectively distinguish the non-anomalous spectrogram from the anomalous spectrogram. Hence, PredNet achieves small AUC. The conventional approach also achieves poor AUC. In contrast to both these schemes, our proposed WANDA first extracts features containing the anomaly signal information, and removes all the unnecessary background noise from the spectrogram. This gives the anomaly detection engines a set of highly distinguishable features to differentiate the known and unknown signals for all SIR values. Thus, WANDA achieves substantially higher AUC scores than both benchmark schemes.

6.B   Advantages of H-score Based RF Signature Extraction.
Figs. 3 and 4 plot the ROC curves of the proposed and benchmark schemes for different SIR values. In the subplots 3(b), 3(d), 4(b), and
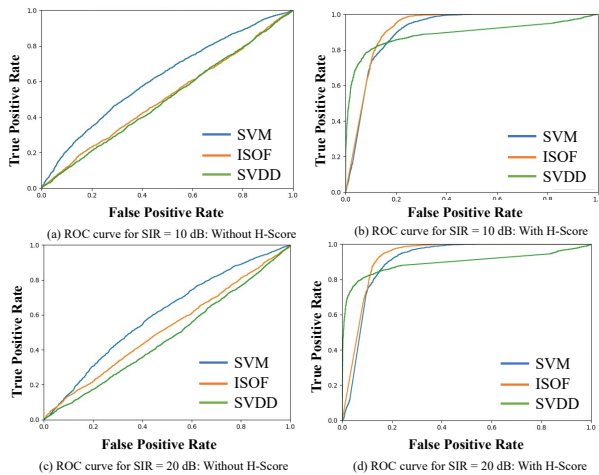
5

(a) ROC curve for SIR = 10 dB: Without H-Score

(b) ROC curve for SIR = 10 dB: With H-Score

(c) ROC curve for SIR = 20 dB: Without H-Score

(d) ROC curve for SIR = 20 dB: With H-Score

Figure 4: ROC Curves of Comparison for SIRs 10 and 20 dB.

4(d), "With H-Score" represent ROC curves of anomaly detection scheme equipped with H-score CNN based feature extraction approach. Within these figures, the blue colored curve represents ROC of our proposed WANDA framework, where the orange and green colored curve represents ROC of the H-score CNN with ISOF and SVDD based anomaly detection engines, respectively. Meanwhile, in figures 3(a), 3(c), 4(a), and 4(d)), "without H-Score" represent ROC curves of the conventional anomaly detection schemes, i.e., without H-score CNN based feature extraction approach.

We emphasize that ROC curve of a particular anomaly detection approach provides a trade-off between the TPR and FPR. For the improved RF anomaly detection, a given scheme needs to provide high TPR at the cost of low FPR. Figs. 3 and 4 depict that regardless of SIR values, the TPR and FPR of the conventional anomaly detection schemes remain the same, i.e., these schemes cannot achieve high TPR and low FPR, simultaneously. In other words, these scheme cannot efficient differentiate between signal classes with and without anomaly. In contrast, the H-score CNN equipped RF anomaly detection approach achieves notably improved TPR-FPR trade-off for all the SIR values. For instance, at SIR 0 dB, the standalone OC-SVM approach achieves 30% TPR at the cost of 40% FPR. However, when H-score CNN based feature extraction approach is augmented before the OC-SVM classifier (i.e., WANDA), 80% TPR is achieved at the cost of only 20% FPR. The aforementioned results confirm the fact that the proposed H-score CNN based feature extraction approach can provide useful information to RF anomaly detection engine, leading to notably improved trade-off between the TPR and FPR for RF anomaly detection. Fig. 3 and 4 also illustrate that in the presence of an H-score CNN based feature extraction approach, all the anomaly detection engines (OC-SVM, ISOF, and SVDD) achieves almost the same TPR-FPR trade-off for high SIRs (i.e., SIR greater than 0 dB), whereas at -10 dB SIR OC-SVM achieves much improved TPR-FPR trade-off compared to both ISOF and SVDD based anomaly detection engines. Because of the superior TPR-FPR trade-off at all SIR values, we select OC-SVM in the anomaly detection engine of the proposed WANDA framework.

7 Conclusion

In this paper, an automated and fully unsupervised LPI RF anomaly signal detection framework called WANDA was proposed. Unlike the traditional anomaly signal detection approaches, prior to anomaly detection, WANDA extracts the low-dimensional informative features from the high-dimensional spectrogram by eliminating noise and without using any manually defined labels. Such a capability of extracting features enables WANDA to detect RF anomaly signals at both high and low SIR values. WANDA comprises an H-score CNN-aided feature extraction step, followed by an OC-SVM-based anomaly detection step. By employing this two-step approach, WANDA optimized the detection process and ensures an accurate identification of RF anomalies. For a case study, we applied the proposed WANDA framework to detect unknown Bluetooth signals within a WiFi spectrum, using a realistic CRAWDAD dataset. Our experiments demonstrated that WANDA significantly outperforms both state-of-the-art PredNet and conventional anomaly detection schemes in terms of AUC score and TPR-FPR trade-off across a wide range of SIR values.

References

[1] H. Bouzabia, T. N. Do, and G. Kaddoum, "Deep learning-enabled deceptive jammer detection for low probability of intercept communications," *IEEE Systems Journal*, vol. 17, pp. 2166–2177, June 2023.

[2] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Malicious user detection in a cognitive radio cooperative sensing system," *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2488–2497, 2010.

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

[4] S. Yin, S. Li, and J. Yin, "Temporal-spectral data mining in anomaly detection for spectrum monitoring," in *2009 5th International Conference on Wireless Communications, Networking and Mobile Computing*, pp. 1–5, 2009.

[5] K. Tekbiyik, O. Akbunar, A. R. Ekti, A. Gorçin, and G. K. Kurt, "Multi–dimensional wireless signal identification based on support vector machines," *IEEE Access*, vol. 7, pp. 138890 – 138903, 2019.

[6] Q. Feng, Z. Dou, C. Li, and G. Si, "Anomaly detection of spectrum in wireless communication via deep autoencoder," in *Advances in Computer Science and Ubiquitous Computing* (J. J. J. H. Park, Y. Pan, G. Yi, and V. Loia, eds.), (Singapore), pp. 259–265, Springer Singapore, 2017.

[7] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, "Unsupervised wireless spectrum anomaly detection with interpretable features," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 637–647, 2019.

[8] T. J. O'Shea, T. C. Clancy, and R. W. McGwier, "Recurrent neural radio anomaly detection," 2016.

[9] Z. Li, Z. Xiao, B. Wang, B. Y. Zhao, and H. Zheng, "Scaling deep learning models for spectrum anomaly detection," *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 2019.

[10] N. Tandiya, A. Jauhar, V. Marojevic, and J. H. Reed, "Deep predictive coding neural network for rf anomaly detection in wireless networks," *IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2018.

[11] C. P. Robinson, D. Uvaydov, S. D'Oro, and T. Melodia, "Narrowband interference detection via deep learning," *IEEE International Conference on Communications*, pp. 1–6, 2023.

[12] D. Roy, V. Chaudhury, C. Tassie, C. Spooner, and K. R. Chowdhury, "Icarus: Learning on iq and cycle frequencies for detecting anomalous rf underlay signals," *IEEE INFOCOM 2023*, pp. 1–9, 2023.

[13] D. A. Ravi, "Identifying and prioritizing critical information in military iot: Video game demonstration," *Phd Thesis, Virginia Tech*, 2022.

[14] S.-L. Huang, A. Makur, G. W. Wornell, and L. Zheng, "On universal features for high-dimensional learning and inference," 2019.

[15] S.-L. Huang, L. Zhang, and L. Zheng, "An information-theoretic approach to unsupervised feature selection for high-dimensional data," in *2017 IEEE Information Theory Workshop (ITW)*, pp. 434–438, 2017.

[16] L. Wang, J. Wu, S.-L. Huang, L. Zheng, X. Xu, L. Zhang, and J. Huang, "An efficient approach to informative feature extraction from multimodal data," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5281–5288, Jul. 2019.

[17] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," 2019.

[18] M. Schmidt, D. Block, and U. Meier, "CRAWDAD dataset owl/interference (v. 2019-02-12)." Downloaded from https://crawdad.org/owl/interference/20190212, Feb. 2019.