RF ONE-CLASS CLASSIFICATION VIA SHIFTED DEEPSVDD

Edward Reehorst

Air Force Research Laboratory
Sensors Directorate
Wright-Patterson AFB, OH, USA

Philip Schniter

The Ohio State University
Dept. of Elec. and Comp. Eng.
Columbus, OH, USA

ABSTRACT

In this work, we develop a novel method for one-class classification of radio frequency (RF) data. One-class classification (also known as novelty detection, out-of-distribution detection, or unsupervised anomaly detection) is the machine learning problem where one is given a dataset of inlier samples and the goal is to detect outlier samples, or novelties, without any examples of those novelties. We propose to extend DeepSVDD, as proposed by Ruff et al. in 2018, by incorporating "shifting transforms" made popular in the contrastive learning literature with a recent GLRT-based score-combining technique. In experiments with the RadioML 2018 dataset, we find that the proposed "Shifted DeepSVDD" outperforms the original DeepSVDD as well as traditional one-class support vector machine (SVM) methods.

1. INTRODUCTION

In the context of radio frequency (RF) modulation classification, we consider the one-class (OC) classification problem, where one is given examples of an inlier set and the goal is to detect outliers, for which no examples are given. This problem arises, e.g., in spectrum monitoring, where the modulation types approved for transmission are known and the objective is to detect non-approved transmissions. In this case, we know nothing specific about the non-approved transmissions except that they don't use an approved modulation. Thus, the detector is trained only on inlier samples.

OC classification is also known as novelty detection, out-of-distribution detection, and unsupervised anomaly detection. For low-dimensional data, there exist many well-known classical approaches based on density estimation, e.g., using Mahalanobis distance [1], Gaussian mixture models [2], nearest-neighbor distance [3, 4], or kernel density estimation [5]. However, such methods typically don't perform well on high-dimensional data, such as images or RF signals, due to the curse of dimensionality [6]. There also exist classical approaches based on kernel PCA [7] and support vector machines (SVMs), such as the OCSVM [8] and the support-vector data description (SVDD) method from [9], but selecting kernels and hand-crafting relevant features can be challenging, since, in the raw feature space, novelties can be very close to inliers according to traditional ℓ_p and kernel-based metrics.

For OC classification of images, many deep neural network (DNN)-based techniques have been developed, such as deep autoencoders, deep SVMs, deep generative networks (e.g., generative adversarial networks and normalizing flows), and deep self-supervised methods, as described in recent surveys [10–13]. Two that are particularly relevant to our work include the DeepSVDD approach by Ruff et al. [14] and the "shifting transform"-based approaches by Tack et al. [15] and Sohn et al. [16]. In our experience, however, very few of these image-based OC classification methods perform acceptably on RF data.

For OC classification of RF data, the literature is much more sparse. In the special case that the inliers come with class labels, the problem becomes "open set" classification, for which a number of deep-network RF techniques have been proposed [17–19]. But for the more general OC problem, the existing approaches are based on adaptations of the OCSVM [20, 21].

In this work, we propose a new approach to deep OC classification that combines DeepSVDD [14] with the shifting-based methods [15, 16] and the recent generalized likelihood ratio test (GLRT)-based score-combining technique from [22]. Our experiments, performed on the RadioML 2018 dataset [23], demonstrate that our proposed "Shifted DeepSVDD" method outperforms standard DeepSVDD and traditional baselines.

2. BACKGROUND

2.1. DeepSVDD

DeepSVDD [14] is a deep learning based method for OC classification, where a DNN is trained to produce feature representations that are close to the mean feature vector

for the inlier training samples. This is done by minimizing the loss

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_i) - \boldsymbol{c}\|^2 + \frac{\lambda_0}{2}\sum_{k=1}^{L} \|\boldsymbol{W}_{\boldsymbol{\theta}}^k\|_F^2, \quad (1)$$

where $\{\boldsymbol{x}_i\}_{i=1}^n$ are inlier training samples in $\mathcal{X}$, $\boldsymbol{f}_{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{R}^d$ is a DNN that outputs a d-dimensional feature vector, $\boldsymbol{W}_{\boldsymbol{\theta}}^k$ is the DNN weight matrix from layer k out of L total layers, $\|\cdot\|_F$ is the Frobenius norm, and $\boldsymbol{c} \in \mathbb{R}^d$ is the mean or center vector, defined as

$$\boldsymbol{c} = \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{f}_{\boldsymbol{\theta}_0}(\boldsymbol{x}_i)\right]_{|\cdot|>\epsilon}. \quad (2)$$

In (2), $\boldsymbol{\theta}_0$ are the DNN parameters at initialization and $[\cdot]_{|\cdot|>\epsilon}$ denotes the element-wise projection ϵ-away from zero in magnitude. For example, $[0.1]_{|\cdot|>1} = 1$ and $[-0.1]_{|\cdot|>1} = -1$. Once $\boldsymbol{c}$ is determined, it is not updated during training, which stabilizes training and avoids trivial solutions.

Once trained, the DeepSVDD score

$$s(\boldsymbol{x}) = -\|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{c}\|, \quad (3)$$

is used to asses a new test sample $\boldsymbol{x}$, where lower scores suggest novelties and higher scores suggest inliers. To make a decision on whether $\boldsymbol{x}$ is a novelty or inlier, the score is compared to a threshold.

2.2. Shifting Transforms

Shifting transforms were popularized in the literature on contrastive OC image classification [15, 16]. In contrast to data-augmentation image transforms like cropping, horizontal flipping, and color jittering, where the goal is to create new samples that are semantically *identical* to the original sample, shifting transforms aim to generate new samples that are semantically *distinct* from the original sample, yet not completely unrelated. One of the most common shifting transforms for images is {0,90,180,270}-degree rotation. Shifting transforms have been shown to benefit contrastive OC classification by making the inlier feature representation space less crowded [16] and by providing multiple evaluations of the scoring function that can be combined to improve the robustness of OC classification [15].

2.3. Score Combining

For their OC image classifier, Tack et al. [15] combined the shift-based scores $\{s_j(\cdot)\}_{j=1}^c$ using simple mean, i.e., $\frac{1}{c}\sum_{j=1}^c s_j(\cdot)$. Reehorst and Schniter [22] proposed a more sophisticated GLRT-based score-combining approach for OC classification and demonstrated improved performance with image data. We now briefly review the approach from [22].

Suppose that one is given c inlier score functions $\{s_j(\cdot)\}_{j=1}^c$. By "inlier" scores, we mean that the scores are designed to produce larger values with inliers than novelties. As a first step, each $s_j(\cdot)$ is transformed to a $z_j(\cdot)$ for which the output distribution is approximately Gaussian when evaluated on the inlier data. In particular, $z_j(\cdot)$ is constructed using the quantile transform, i.e.,

$$z_j(\boldsymbol{x}) \triangleq \Phi^{-1}(\widehat{F}_j(s_j(\boldsymbol{x}))), \quad (4)$$

where $\widehat{F}_j(\cdot)$ is the empirical cumulative distribution function (CDF) of $\{s_j(\boldsymbol{x}_i)\}$ evaluated on a hold-out set of the inliers $\{\boldsymbol{x}_i\}$ and $\Phi^{-1}(\cdot)$ is the inverse CDF of the standard normal distribution. Thus the inlier "z-scores" $\{z_j(\boldsymbol{x}_i)\}_i$ are approximately standard-normal distributed. Because i) the $s_j(\cdot)$ are designed so that novelties score lower than inliers, ii) the quantile transform is monotonic, and iii) the inlier z-scores are zero-mean, we can conclude that the z-scores for novelties will be negative on average.

For tractability, we will assume that the combined z-scores $\boldsymbol{z} = [z_1, \ldots, z_c]^\mathsf{T}$ are Gaussian distributed with identity covariance for both inliers and novelties. This allows us to formulate OC classification as a composite hypothesis test

$$\begin{aligned} H_0 :&\quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \\ H_1 :&\quad \boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\alpha}, \boldsymbol{I}),\ \ \boldsymbol{\alpha} \in \mathbb{R}_-^c, \end{aligned} \quad (5)$$

where $\boldsymbol{\alpha}$ is the unknown mean of the novelty distribution, $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate normal with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and $\mathbb{R}_-^c$ is set of c-length real vectors with negative elements.

A common way to tackle composite hypothesis tests is via the generalized likelihood ratio test (GLRT) [24], which in this case manifests as

$$\mathrm{GLR}(\boldsymbol{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau' \ \text{ for } \ \mathrm{GLR}(\boldsymbol{z}) \triangleq \frac{\max_{\boldsymbol{\alpha}\in\mathbb{R}_-^c} \mathcal{N}(\boldsymbol{z}; \boldsymbol{\alpha}, \boldsymbol{I})}{\mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{I})} \quad (6)$$

and some user-chosen $\tau' > 0$, or equivalently as

$$t_{\mathrm{glrt}}(\boldsymbol{z}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau \ \text{ for } \ t_{\mathrm{glrt}}(\boldsymbol{z}) \triangleq \max_{\boldsymbol{\alpha}\in\mathbb{R}_-^c} \log \frac{\mathcal{N}(\boldsymbol{z}; \boldsymbol{\alpha}, \boldsymbol{I})}{\mathcal{N}(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{I})} \quad (7)$$

for some $\tau \in \mathbb{R}$. It is shown in [22] that

$$t_{\mathrm{glrt}}(\boldsymbol{z}) = (\tfrac{1}{2}\boldsymbol{z}^- - \boldsymbol{z})^\mathsf{T}\boldsymbol{z}^-, \quad (8)$$

where $\boldsymbol{z}^-$ is the projection of $\boldsymbol{z}$ onto the negative hyper-octant. That is, the jth element of $\boldsymbol{z}^-$ is given by

$$z_j^- \triangleq \min\{z_j, 0\}. \quad (9)$$

Thus (8) provides a simple and principled way of combining inlier scores for OC classification.

3. PROPOSED METHOD

We propose to extend the DeepSVDD approach described in Section 2.1 to use $c > 1$ shifting transforms, as described in Section 2.2. We then propose to combine the resulting c inlier scores using the GLRT-based score-combining method described in Section 2.3. Our "Shifted DeepSVDD" method is detailed below.

Given shifting transformations $\boldsymbol{T}_j : \mathcal{X} \to \mathcal{X}$ for $j = 1, \ldots, c$, we first compute each center vector

$$\boldsymbol{c}_j = \left[\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{f}_{\boldsymbol{\theta}_0}(\boldsymbol{T}_j(\boldsymbol{x}_i)) \right]_{|\cdot| > \epsilon} \tag{10}$$

from the inlier data $\{\boldsymbol{x}_i\}_{i=1}^n$ using randomly initialized DNN weights $\boldsymbol{\theta}_0$. The training loss is then defined as

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{j=1}^{c} \frac{\lambda_j}{n} \sum_{i=1}^{n} \|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{T}_j(\boldsymbol{x}_i)) - \boldsymbol{c}_j\|^2 + \frac{\lambda_0}{2} \sum_{k=1}^{L} \|\boldsymbol{W}_{\boldsymbol{\theta}}^k\|_F^2 \tag{11}$$

with hyperparameters $\{\lambda_j\}_{j=0}^c$. The network weights $\boldsymbol{\theta}$ are then trained to minimize $\mathcal{L}(\boldsymbol{\theta})$ using mini-batch stochastic gradient descent (SGD). Once trained, the DNN provides an inlier score

$$s_j(\boldsymbol{x}) = -\|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{T}_j(\boldsymbol{x})) - \boldsymbol{c}_j\| \tag{12}$$

for each shift j. Finally, the c inlier scores are combined using the GLRT-based technique in Section 2.3.

To apply our technique to RF waveform data, where $\boldsymbol{x} \in \mathbb{C}^m$ for some integer m, we need a suitable architecture for $\boldsymbol{f}_{\boldsymbol{\theta}}(\cdot)$. Starting with the standard image-based DeepSVDD architecture, we replace the 2D convolutions with 1D complex convolutions and apply ReLU and batch norm separately to the real and imaginary channels, as recommended in [25].

After experimenting with various choices of shifting transforms, we settled on the $c = 2$ transforms

$$\boldsymbol{T}_1(\boldsymbol{x}) = \boldsymbol{x} \tag{13}$$

$$\boldsymbol{T}_2(\boldsymbol{x}) = \boldsymbol{x} \odot \exp(\mathrm{j}2\pi f_{\mathrm{mod}} \boldsymbol{t}), \tag{14}$$

where $\odot$ denotes element-wise complex multiplication, $\mathrm{j} = \sqrt{-1}$, $f_{\mathrm{mod}} \in [-0.5, 0.5]$, and $\boldsymbol{t} \triangleq [0, 1, \ldots, m-1]^\mathsf{T}$. Note that $\boldsymbol{T}_1(\cdot)$ is an identity transform and $\boldsymbol{T}_2(\cdot)$ shifts the frequency by f_{mod} radians per sample. In our experiments, we set $f_{\mathrm{mod}} = 0.2$.

4. EXPERIMENTS

Our experiments used the RadioML 2018 dataset [23]. The dataset contains 24 modulation classes, each containing 106 496 RF waveforms of length $m = 1024$ spanning SNRs between -20 and 30 dB. Five different OC classification problems were created. In each, one modulation class forms the inliers while the union of the others forms the novelties. The inlier classes where OOK, 32PSK, 16QAM, AM-DSB-WC, and FM, respectively, which include a mix of phase modulation, amplitude modulation, and frequency modulation. For each inlier class, we partitioned the data into training and test sets. Once trained using the inlier data (since the methods had no access to novelty data), the models were used to classify test samples from both the inlier and novelty classes. Classification performance is reported as AUC, i.e., area under the receiver operating curve.

We compare our Shifted DeepSVDD method to standard non-shifted DeepSVDD (using otherwise identical settings) and two versions of the OCSVM [8], one of which operates on raw RF signals $\boldsymbol{x}$ and the other of which operates on auto-correlation features (ACF) that are computed from $\boldsymbol{x}$ as described in [20, 21]. Additionally, we evaluate Shifted DeepSVDD with both the GLRT-based score-combining method from Section 2.2 and the mean-based score-combining used in [15]:

$$t_{\mathrm{mean}}(\boldsymbol{x}) \triangleq \frac{1}{c} \sum_{j=1}^{c} s_j(\boldsymbol{x}). \tag{15}$$

All DeepSVDD implementations use convolution kernels of length 5; $\epsilon = 0.1$ in (10); $\lambda_0 = 10^{-6}$, $\lambda_1 = 1$, and $\lambda_2 = 0.1$ in (11); and the Adam optimizer [26] when minimizing (11).

Due to the inherent randomness in the initialization $\boldsymbol{\theta}_0$ that defines $\boldsymbol{c}_j$ and in mini-batch SGD, we trained 10 versions of each DeepSVDD and Shifted DeepSVDD model, and we report the mean and standard error of the AUC across those 10 trials.

Results from these experiments are shown in Table 1. There we see that the proposed GLRT-based Shifted DeepSVDD method outperformed the competitors on four of the five inlier classes, as well as on average. Furthermore, the proposed method outperforms standard (non-shifted) DeepSVDD [14] on all five inlier classes. In addition, the GLRT-based combining scheme outperforms the mean-based scheme [15] for all but one inlier class, where they tie. Finally, the DNN-based schemes outperform the OCSVM methods in most cases, as well as on average. Altogether, these results suggest i) that DeepSVDD-based techniques outperform classical OCSVM-based techniques for RF waveform OC classification, ii) that shifting-based transforms [15, 16] benefit DeepSVDD, and iii) that GLRT-combining [22] outperforms simple mean combining.

We now look more closely at the experiment with AM-DSB-WC inliers, which was the only one where OCSVM performed best. For this experiment, one of

Method	Inlier Class					Avg.
	OOK	32PSK	16QAM	AM-DSB-WC	FM	
OCSVM – raw RF	0.719	0.500	0.486	0.790	0.794	0.658
OCSVM – ACF	0.722	0.623	0.600	**0.877**	0.674	0.700
DeepSVDD	0.677 ± 0.082	0.710 ± 0.019	0.673 ± 0.012	0.827 ± 0.007	0.926 ± 0.011	0.763
Shifted DeepSVDD – mean combine	0.827 ± 0.007	0.769 ± 0.052	**0.697** ± 0.016	0.682 ± 0.020	0.816 ± 0.032	0.758
Shifted DeepSVDD – GLRT (ours)	**0.844** ± 0.008	**0.801** ± 0.029	**0.697** ± 0.011	0.842 ± 0.012	**0.944** ± 0.005	**0.826**

Table 1: AUC for OC classification of RadioML 2018 waveforms. For DNN-based methods, the table reports the average ± standard error across 10 training runs.

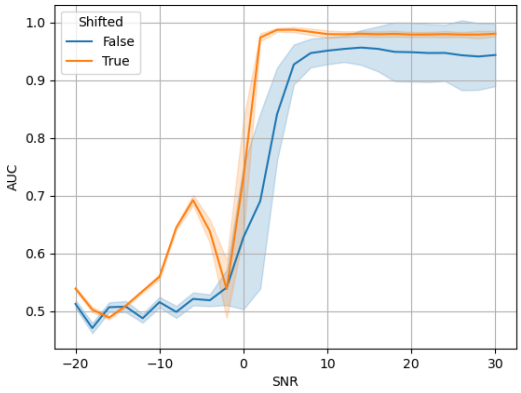


Fig. 1: AUC versus SNR for OOK inliers.

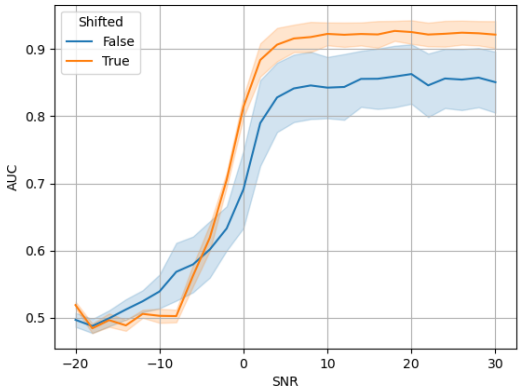


Fig. 2: AUC versus SNR for 32PSK inliers.

the two scores used by the Shifted DeepSVDD was very weak, which is evident by the very poor performance of mean-combined Shifted DeepSVDD. Although the GLRT-combined method did not win this contest, it came in second best. This illustrates the power of GLRT score-combining: it was able to combine the strong score (i.e., the one used by standard DeepSVDD) with the weak score in a way that performed better than the strong score alone or the mean of the two scores.

Figures 1-5 plot AUC versus SNR for the proposed GLRT-combined shifted DeepSVDD and standard (non-shifted) DeepSVDD for each of the five inlier classes under test. In each plot, the solid traces show mean AUC over 10 training runs and the shaded region shows the standard deviation. As expected, AUC tends to increase with SNR. The proposed Shifted-DeepSVDD gives higher AUC than standard DeepSVDD in nearly every case, with the exception of 16QAM in Fig. 3.[1]

[1]To reconcile the results of Table 1 with Fig. 3, we remind the reader that AUC on the full dataset is not equal to the average of AUCs on different subsets (in this case, at different SNRs). For example, consider a 4-sample dataset where the outliers have scores 0 & 5 and the inliers have scores 1 & 6. If we form one subset with scores [0,1] and another subset with scores [5,6], then each subset would have a perfect AUC of 1, while the AUC on the full dataset is <1.

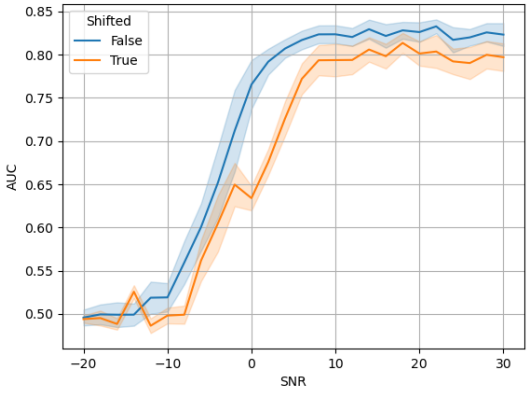


Fig. 3: AUC versus SNR for 16QAM inliers.

5. CONCLUSION

With the goal of one-class classification of RF waveforms, this paper proposed a new method called Shifted DeepSVDD. In particular, the proposed method incorporates shifting transforms, as popularized in the contrastive-learning literature [15, 16], into DeepSVDD [14] and combines the resulting scoring functions using the GLRT-based technique from [22]. The pro-

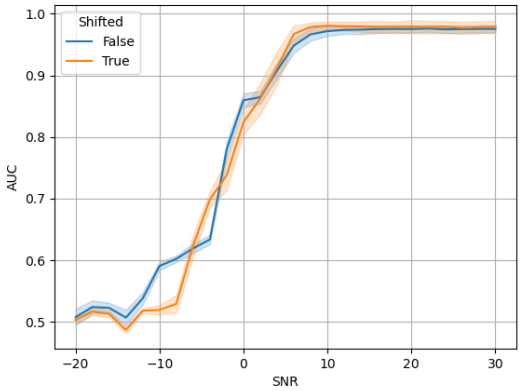


Fig. 4: AUC versus SNR for AM-DSB-WC inliers.

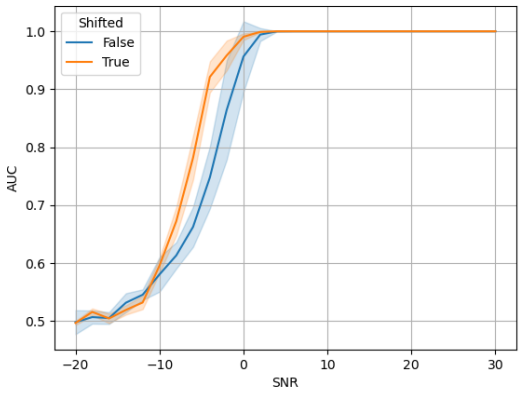


Fig. 5: AUC versus SNR for FM inliers.

posed technique was evaluated using RadioML 2018 waveforms and shown to outperform OCSVM (with raw waveforms or ACF features), standard DeepSVDD, and mean-combined Shifted DeepSVDD. Future work might consider applying Shifted DeepSVDD to other data modalities using appropriate choices of shifting transforms and DNN architecture.

6. REFERENCES

[1] J. Laurikkala, M. Juhola, E. Kentala, N. Lavrac, S. Miksch, and B. Kavsek, "Informal identification of outliers in medical data," in *Intl. Workshop Intell. Data Anal. Med. Pharma.*, pp. 20–24, 2000.

[2] C. M. Bishop, "Novelty detection and neural network validation," *IEE Proc. V: Vision, Image & Signal Process.*, vol. 141, no. 4, pp. 217–222, 1994.

[3] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," in *Proc. ACM SIGMOD Intl. Conf. Manag. Data*, pp. 427–438, 2000.

[4] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Intl. Conf. Manag. Data*, pp. 93–104, 2000.

[5] L. J. Latecki, A. Lazarevic, and D. Pokrajac, "Outlier detection with kernel density functions," in *Proc. Mach. Learn. Data Mining Pattern Recog.*, vol. 7, pp. 61–75, 2007.

[6] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. and Data Mining*, vol. 5, no. 5, pp. 363–387, 2012.

[7] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863–874, 2007.

[8] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, pp. 1443–1471, July 2001.

[9] D. Tax and R. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, pp. 45–66, 2004.

[10] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," *arXiv:1901.03407*, 2019.

[11] F. Di Mattia, P. Galeone, M. De Simoni, and E. Ghelfi, "A survey on GANs for anomaly detection," *arXiv:1906.11632*, 2019.

[12] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surveys*, vol. 54, no. 2, pp. 1–38, 2021.

[13] L. Ruff, J. R. Kauffmann, and R. A. Vandermeulen, "A unifying review of deep and shallow anomaly detection," *Proc. IEEE*, vol. 109, no. 5, pp. 756–795, 2021.

[14] L. Ruff, R. Vandermeulen, N. Görnitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Müller, and M. Kloft, "Deep one-class classification," in *Proc. Intl. Conf. Mach. Learn.*, 2018.

[15] J. Tack, S. Mo, J. Jeong, and J. Shin, "CSI: Novelty detection via contrastive learning on distributionally shifted instances," in *Proc. Neural Info. Process. Syst. Conf.*, pp. 11839–11852, 2020.

[16] K. Sohn, C.-L. Li, J. Yoon, M. Jin, and T. Pfister, "Learning and evaluating representations for deep one-class classification," in *Proc. Intl. Conf. Learn. Rep.*, 2021.

[17] R. V. Chakravarthy, H. Liu, and A. M. Pavy, "Open-set radar waveform classification: Comparison of different features and classifiers," in *Radiology*, pp. 542–547, 2020.

[18] S. Hanna, S. Karunaratne, and D. Cabric, "Open set wireless transmitter authorization: Deep learning approaches and dataset considerations," *IEEE Trans. on Cognitive Communications and Networking*, 2021.

[19] X. Zhang, T. Li, P. Gong, R. Liu, Z. Zha, and W. Tang, "Open set recognition of communication signal modulation based on deep learning," *IEEE Communications Letters*, 2022.

[20] B. Rigling and C. Roush, "ACF-Based classification of phase modulated waveforms," in *Radiology*, pp. 287–291, 2010.

[21] A. Pavy and B. Rigling, "Phase modulated radar waveform classification using quantile one-class SVMs," in *Radiology*, pp. 745–750, 2015.

[22] E. Reehorst and P. Schniter, "Score combining for contrastive OOD detection," *arXiv:2501.12204*, 2025.

[23] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, 2018.

[24] E. L. Lehmann, J. P. Romano, and G. Casella, *Testing Statistical Hypotheses*, vol. 3. Springer, 2005.

[25] M. Wharton, A. Pavy, and P. Schniter, "Deep neural networks for radar waveform classification," in *Proc. Asilomar Conf. Signals Syst. Comput.*, pp. 1623–1627, 2021.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Intl. Conf. Learn. Rep.*, 2015.