Deep Learning Approaches for Open Set Wireless Transmitter Authorization

Samer Hanna, Samurdhi Karunaratne, and Danijela Cabric
Electrical and Computer Engineering Department,
University of California, Los Angeles
samerhanna@ucla.edu, samurdhi@ucla.edu, danijela@ee.ucla.edu

***Abstract*—Wireless signals contain transmitter specific features, which can be used to verify the identity of transmitters and assist in implementing an authentication and authorization system. Most recently, there has been a wide interest in using deep learning for transmitter identification. However, the existing deep learning work has posed the problem as closed set classification, where a neural network classifies among a finite set of known transmitters. No matter how large this set is, it will not include all transmitters that exist. Malicious transmitters outside this closed set, once within communications range, can jeopardize the system security. In this paper, we propose a deep learning approach for transmitter authorization based on open set recognition. Our proposed approach identifies a set of authorized transmitters, while rejecting any other unseen transmitters by recognizing their signals as outliers. We propose three approaches for this problem and show their ability to reject signals from unauthorized transmitters on a dataset of WiFi captures. We consider the structure of training data needed, and we show that the accuracy improves by having signals from known unauthorized transmitters in the training set.**

***Index Terms*—Transmitter Identification, Deep Learning, Open set recognition, authorization, physical layer authentication**

I. Introduction

With the growth in the number of wirelessly connected devices, securing them has become more challenging. Part of securing wireless devices is authentication; the process of verifying their identity. While there exist many cryptography based methods for authentication, they are not suitable for many internet-of-things devices that have limited computation and power budget.

Physical layer authentication (PLA) enables devices to be authenticated without having to decode the data and typically without requiring additional signaling overhead [1]. Active PLA overlays a tag for authentication over the message thus requiring changes to the physical layer of the transmitters. Passive PLA, on the other hand, uses the channel state information and the transmitter fingerprint due to hardware imperfections to identify transmitters [2], requiring no change to transmitter signals, and hence is more practical.

Approaches for passive PLA either use a set of handcrafted features or deep learning on raw IQ samples. For feature-based PLA, existing works have considered using transmitter fingerprints due to hardware imperfections [3] or channel state information (CSI) [4]. Learning approaches based on handcrafted features rejecting new transmitters have used Gaussian mixture models [4]–[7]. However, the performance of these approaches depends on the receiver quality [8] and requires manual feature engineering.

In contrast, deep learning approaches are more robust and can extract better features from signals, hence leading to higher accuracy compared to feature-based approaches [9]. The existing work in the literature has considered the effect of data representation, neural network architecture, and the wireless channel on the classification accuracy [9]–[17]. Examples of data representations include raw IQ samples [9]–[11], [15], Fourier transform [12], [16], and Wavelet transform [12], [17]. The robustness of the learned features in different channels has also been considered [10]. Network architectures evaluated include DNN [17], CNN [9], [16], [17], RNN [14], and complex neural networks [13]. The main limitation of this body of work is its focus on classification among a closed set of known transmitters. Any transmitter outside this set will be misclassified, hence, jeopardizing the system security. The problem of classifying among known classes and rejecting samples from new classes is known as open set recognition [18]. Many approaches have been proposed to address it in other domains like image classification.

In this paper, we pose the problem of rejecting signals from any transmitter outside a known authorized set as an open set recognition problem. Since this problem was studied extensively, instead of reinventing the wheel, we aim to adapt and evaluate well-established approaches for transmitter authorization. Since the number of authorized transmitters is a system requirement that can vary significantly, we study how these approaches scale in terms of performance and network size with it. To further improve the performance, we propose using a set of known unauthorized transmitters and demonstrate its effectiveness in improving outlier detection. Our results show an average accuracy of 96.8% when separating signals of 10 authorized transmitters and 30 unseen transmitters when using a set of 25 known outliers for training using a WiFi capture.

The problem of open set recognition is more challenging than closed set classification. A closed set classifier determines boundaries that separate the classes it has seen, as shown with the solid blue line in Fig. 1. But, given data from new classes (new unauthorized transmitters), the classifier will predict the nearest class, which poses a security risk for an authentication

This work was supported in part by the CONIX Research Center, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.

978-1-7281-5478-7/20/$31.00 ©2020 IEEE

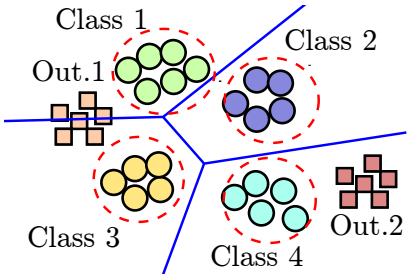


Fig. 1: Known classes are depicted as circles and unknown classes as squares. Solid lines and dashed circles represent classification boundaries and open set recognition, respectively.

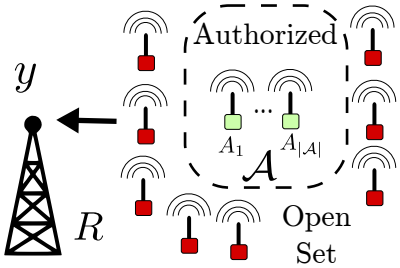


Fig. 2: Signal y is received by receiver R. We want to determine if it was sent by an authorized transmitter in the set $\mathcal{A}$ or a new unseen transmitter.

system. On the other hand, open set classification creates boundaries around the seen distribution, as illustrated with the red dashed circles in Fig. 1, for rejecting samples from new classes. Unlike feature based approaches [4]–[7] which use well separated features like CSI, our approach needs to learn the features that separate authorized transmitters from transmitters for which no training data is available.

The remainder of the paper is organized as follows: we start by formulating the problem in Section II. Section III discuses the considered machine learning approaches. The dataset, the architecture, and the results are presented in Section IV. Section V concludes the paper.

II. System Model and Problem Formulation

We consider a finite set of authorized transmitters given by $\mathcal{A} = \{A_1, A_2, \cdots, A_{|\mathcal{A}|}\}$ that are authorized to send data to a receiver R, where $|\mathcal{A}|$ is the size of the set $\mathcal{A}$. When a transmitter T sends a set of symbols x, the signal received is $f_T(x)$. The function f_T models the transmitter fingerprint determined by the variability of its circuit and also includes the effects of the channel. The authorization problem can be formulated as shown in Fig. 2: receiver R receives a signal y from some transmitter T and wants to determine whether the transmitter T belongs to the authorized set or not, based on y. This can be formulated as the following hypothesis testing:

$$\begin{aligned} \mathcal{H}_0 &: y = f_T(x), T \in \mathcal{A} \\ \mathcal{H}_1 &: y = f_T(x), T \notin \mathcal{A} \end{aligned} \tag{1}$$

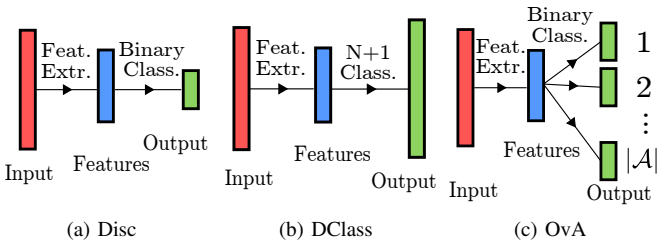


Fig. 3: Architecture of the proposed methods.

Here, $\mathcal{H}_0$ corresponds to an authorized transmitter and $\mathcal{H}_1$ corresponds to an outlier.

Additionally, in cases where each authorized transmitter has different privileges, we might be interested in classifying the transmitter within the authorized set, which can be formulated as finding $\hat{A}$ that maximizes the probability of identifying the true transmitter

$$\hat{A} = \underset{T}{\text{argmax}} \ \Pr(f_T(x) = y \mid y), \qquad T \in \mathcal{A} \tag{2}$$

To improve the outlier detection, we propose using an additional class of known outliers $\mathcal{K} = \{K_1, K_2, \cdots, K_{|\mathcal{K}|}\}$, where $\mathcal{K} \not\subset \mathcal{A}$. Samples from transmitters in $\mathcal{K}$ will be used during training to assist the outlier detector to differentiate between authorized and non-authorized transmitters. But still, the evaluation of any outlier detector is done using a set of unknown outliers $\mathcal{O}$ such that $\mathcal{O} \cap \mathcal{K} = \emptyset$. In practice, samples from the set $\mathcal{K}$ can be obtained by capturing data from a finite number of non-authorized transmitters.

III. Machine Learning Approach

In this section, we discuss the neural network architectures used to solve this problem and the processing performed on the output of these networks to decide if a signal is an outlier. We consider several neural network architectures for outlier detection. These networks consist of a feature extractor followed by one or many classifiers. In terms of training, some of these networks need known outliers to generalize to unseen transmitters, while others don't. We also discuss how the size of $\mathcal{A}$ affects the number of parameters of these networks.

1) Discriminator (Disc): One intuitive approach for outlier detection is to train a discriminator that outputs a decision on whether the signal is an outlier or not. The discriminator has a single scalar output z as shown in Fig. 3a. z is generated by a sigmoid and takes a value between 0 and 1. The labels for authorized transmitters and outliers are $l = 0$ and $l = 1$, respectively. Samples with $l = 1$ used in training are the known outliers, which are necessary for this approach. In the test phase, we declare $\mathcal{H}_1$ if $z > \gamma$ for some threshold γ, else $\mathcal{H}_0$ is declared. In terms of architecture, Disc has the advantage of having a fixed size regardless of $|\mathcal{A}|$. Although this approach does not classify the authorized transmitters, a classifier can be cascaded with a discriminator to achieve this but is not discussed in this work.

2) Discriminating Classifier (DClass): Instead of cascading a discriminator and a classifier, we can directly train a network with $|\mathcal{A}|+1$ outputs, where the additional class corresponds to outliers. Similar to Disc this approache relies on $\mathcal{K}$ for training. As for deciding on outliers, a signal is classified as an outlier if the maximum activation corresponds to the last class, else it is considered authorized without adjustable thresholds. In comparison with Disc, the labels of the transmitters are expected to help DClass learn better features and hence perform better. This comes at the cost of increasing the size of the last layer as $|\mathcal{A}|$ increases.

3) One Vs All (OvA): A simple way to modify Disc to include classification with modifiable thresholds is to use $|\mathcal{A}|$ instances of it; one for each transmitter. However, this method requires a high computational complexity due to having $|\mathcal{A}|$ feature extractors performing the same task. A better approach, proposed in [19], is shown in Fig. 3c. In this approach, we use $|\mathcal{A}|$ binary classifiers, each deciding for a transmitter while sharing the same feature extractor. The output of this network will be a vector $\mathbf{z}$ of $|\mathcal{A}|$ real numbers such that $\mathbf{0} \leq \mathbf{z} \leq \mathbf{1}$, where $\mathbf{0}$ and $\mathbf{1}$ are the vectors of all-zeros and all-ones, respectively. Following the notation in [19], the labels for a sample from authorized transmitter A_i will have $l_i = 1$ and $l_j = 0\ \forall j \neq i$ while known outliers will have labels equal to $\mathbf{0}$. For this architecture, the threshold will be a vector $\boldsymbol{\gamma}$, where element γ_i is the threshold for z_i. The binary classifier i decides that the input sample belongs to class i if $z_i > \gamma_i$; otherwise, it does not belong to class i. We declare the signal to be an outlier (corresponding to $\mathcal{H}_1$), if all discriminators declare the signal to be not within their class ($\mathbf{z} \leq \boldsymbol{\gamma}$), and to be within the authorized set (corresponding to $\mathcal{H}_0$) otherwise.

Note that OvA, unlike DClass and Disc, does not require a known set of outliers, since for samples of any class i, $l_i = 0$ for signals from other classes. OvA, however, requires an entire binary classifier for each authorized transmitter. hence, among the proposed architectures, it has the worse scalability with respect to $|\mathcal{A}|$, in terms of the number of learnable parameters.

Both OvA and Disc have adjustable thresholds. The value of these thresholds determines their sensitivity to outliers. A tight threshold would lead to signals from authorized transmitters being mistakenly rejected (high probability of false alarm P_{FA}) and a loose threshold would fail to recognize many outlier signals (low probability of detection P_D). This trade-off is commonly visualized by the receiver operating characteristic (ROC) showing both P_{FA} and P_D for a specific receiver. In the Appendix, we describe how this trade-off is implemented and state how a specific threshold is chosen to calculate the outlier detection accuracy.

IV. Experimental Evaluation

We start by describing the dataset used and evaluate the performance of the proposed network architectures on the dataset as we change the size and composition of $\mathcal{A}$ and $\mathcal{K}$.

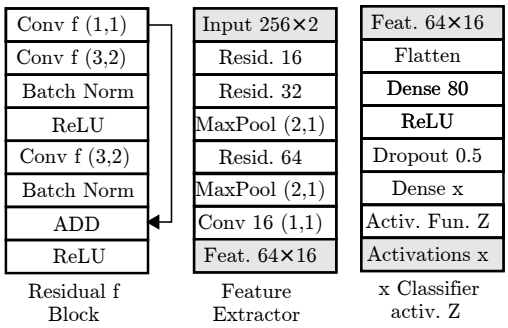


Fig. 4: Detailed architecture of the feature extractor (made of residual blocks with f filters), and a classifier with x outputs.

A. Dataset

The dataset was captured using off-the-shelf WiFi modules (Atheros 5212, 9220, and 9280) as transmitters and a software defined radio (USRP N210) as a receiver, from the Orbit testbed [20]. The choice of the Orbit testbed is made due to the ease of access to many transmitters using realistic hardware while being able to isolate external environmental disturbances. The nodes in Orbit are organized in a 20×20 grid with a separation of one meter; the receiver was chosen near the center and 71 transmitters were randomly chosen, to make many nodes experience similar channels due to the symmetry.

The capture was done over Channel 11 which has a center frequency of 2462 MHz and a bandwidth of 20 MHz. All transmitters were configured to have the same fake MAC address and same IP address to avoid providing any signal based clues about the identity of the transmitter. Captures were taken at a rate of 25 Msps for one second. After the IQ capture was complete, the packets were extracted using energy detection. The number of packets obtained from each transmitter during the capture period varied between 200 and 1500 packets with a mean of 800 packets. This variability is due to WiFi rate control. From each packet, we used the first 256 samples, containing the preamble, without any synchronization or further preprocessing.

B. Network Architectures and Training

We consider the three previously proposed architectures Disc, DClass, and OvA. As stated earlier, these architectures consist of a feature extractor that processes the raw IQ samples and outputs features, followed by a number of classifier blocks. Our focus in this paper is on the approach, not the architecture, so, we used the same feature extractor for all networks. The feature extractor consists of a series of residual blocks with different numbers of filters as shown in Fig. 4. As for the classifier blocks, the architecture for each block is shown in Fig. 4. This network was chosen because similar networks have shown superior performance to CNNs on a similar problem [21]. For Disc, we used one classifier with a sigmoid activation. For OvA, we used N of the classifier blocks with each block similar to that of Disc. As for DClass, we used one

classifier block with $|\mathcal{A}|+1$ outputs and softmax activation. L2 regularization was used in the dense layers with a weight of 0.001 to avoid overfitting.

Note that for OvA and DClass the number of parameters of the neural network increases as the size of $|\mathcal{A}|$ increases. In OvA, for each new authorized transmitter, a new instance of the classifier is added to the architecture with about 80K parameters. For DClass, the size of the last layer increases by 81 parameters for each authorized transmitter. As for Disc, the number of parameters is constant for any $|\mathcal{A}|$.

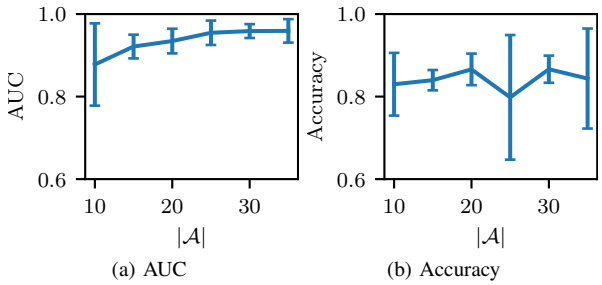
The training was done for 10 epochs using the ADAM optimizer with a learning rate of 0.001. The weights with the lowest validation loss are kept. Samples was first normalized, then augmented by adding noise with a variance of 0.01 and applying a uniformly random phase shift. Cross-Entropy was used as the loss function with classes weighted depending on the number of samples for each class.

C. Transmitter Set Sizes Evaluation

Ideally, we want to train our network using the set of authorized nodes only, regardless of their number. Creating a known outlier set $\mathcal{K}$ would require more transmitters and more data collection. In this section, we explore the effect of changing the size of the authorized set $|\mathcal{A}|$, and the size of the known outlier set $|\mathcal{K}|$, on the ability of the network to distinguish authorized signals from outliers. We start by describing the evaluation metrics and dataset division.

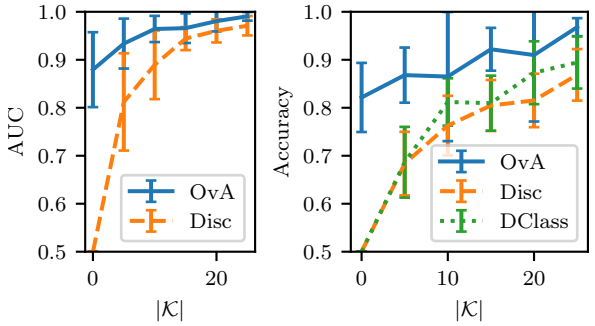
1) Evaluation Metrics and Dataset Division: Since certain subsets of the set of transmitters in our dataset might have more mutually similar signals than others, we try to make our results less specific to a chosen subset of transmitters. To this end, we randomly populate the sets $\mathcal{A}$, $\mathcal{K}$, and $\mathcal{O}$ from the 71 transmitters 10 times in each test we conduct. Results are shown as mean and standards deviation of these 10 realizations. The metrics used for the evaluation of outlier detection are the accuracy and area under the ROC curve (AUC). The accuracy is the percentage of correct predictions calculated over a balanced test set, such that any random or trivial guess would yield 50% accuracy. The area under the ROC curve provides a metric of which model is better on average [22], while the accuracy is what we get for a specific threshold. Although DClass and OvA are capable of classifying signals within the authorized sets, the results of classification were above 99% on the authorized part of the test set, and as classification has been extensively studied in the literature, we omit these results for brevity.

Our training, validation, and test sets are built as follows: for certain values of $|\mathcal{A}|$, $|\mathcal{K}|$, and $|\mathcal{O}|$, we randomly choose transmitters to form our sets $\mathcal{A}$, $\mathcal{K}$, and $\mathcal{O}$. For training and validation, we use 70% of the samples belonging to $\mathcal{A}$, and all the samples belonging $\mathcal{K}$. The shuffled combination of this data is split into 80% for training and 20% for validation. The test set contains all samples from $\mathcal{O}$ and the remaining 30% of $\mathcal{A}$. For different realizations of the sets, the dataset can get highly imbalanced. To avoid degenerate solutions, where the network always predicts the class with the majority of samples, the training loss is weighted.



(a) AUC (b) Accuracy

Fig. 5: Average performance of OvA as we change $|\mathcal{A}|$. Error bars represent the standard deviation.



(a) AUC. (b) Accuracy.

Fig. 6: Average performance of architectures as we change $|\mathcal{K}|$ for $|\mathcal{A}|=10$. Error bars represent the standard deviation.

2) Authorized set: We study how the size of the set $\mathcal{A}$ affects outlier detection by considering OvA with no known outliers, $|\mathcal{K}|=0$. Results are shown in Fig. 5 for $|\mathcal{O}|=30$; we see that as we increase the number of authorized nodes, the AUC increases and its standard deviation decreases. In OvA, to generalize to unseen transmitters, binary classifier i needs to learn the unique features of transmitter i. Seeing signals from more transmitters helps it realize that leading to a better AUC. The accuracy is shown in Fig. 5b, from which we can see that accuracy follows the same trend, except at the point with 25 and 40 authorized transmitters. At these points, the chosen threshold for one realization resulted in a low accuracy, decreasing the mean and increasing the standard deviation. This shows that for specific combinations of authorized and outlier nodes, and a given threshold, we might get lower performance. Since the value of the AUC is high, this shows that another threshold would give a better performance. From these results, we can see that the average accuracy does not go over 90%, while it fluctuates depending on the choice of transmitters for the same method of selecting thresholds.

3) Known set: To further improve performance, we propose using known outliers to help the approaches generalize to unseen transmitters. We evaluate the performance of all architectures as a function of $|\mathcal{K}|$ given $|\mathcal{A}|=10$ and $|\mathcal{O}|=26$. The AUC and accuracy curves are shown in Fig. 6a and

6b respectively. As stated earlier, at $|\mathcal{K}| = 0$, DClass and Disc don't have any outlier samples and predict everything as authorized. From Fig. 6a, we see that the performance of both OvA and Disc improves as we increase the number of known outliers. We note that OvA is performing better. This is explained by recognizing that in OvA each binary classifier sees more samples to reject; the known outliers and the samples not belonging to its class. Thus, it is able to isolate its class better. DClass and Disc, on the other hand, only learn to reject samples from $\mathcal{K}$. Fig 6b follows the same trend with OvA reaching accuracies up to 96% on average. DClass slightly outperforms Disc because the labels of $\mathcal{A}$ help it extract better features compared to Disc. So even if we are not interested in classifying among the nodes in $\mathcal{A}$, including these labels in training improves the outlier detection performance. In the case with $|\mathcal{K}| = 0$, we were not able to attain accuracies above 90% on average, showing that using additional unauthorized transmitters in data collection can lead to significant improvements.

V. Conclusion

In this paper, we formulated transmitter authorization as an open set classification problem where we learn to reject signals from new transmitters not seen during training. We considered three approaches to solve it. OvA does not require known outliers and gives better performance, at the cost of a large increase in the neural network size with the size of the authorized set. DClass and Disc need a large set of known outliers for training to get good performance, with Disc having a slightly lower performance with the advantage of maintaining a constant network size regardless of the number of authorized nodes. In all cases, we have shown that having a set of known outliers improves performance. So far, we have only considered a residual neural network. Further work needs to be carried out to understand the effect of changing the neural network type and architecture for open set recognition.

Appendix

For OvA and Disc, we describe how the threshold used for accuracy is calculated along with the ROC scan.

1) Disc: Ideally, we want the threshold to be as low as possible without falsely rejecting authorized transmitters. This can be done by adapting the threshold to tightly fit the predictions of authorized signals in the training set. We follow the approach proposed in [19], where the predicted output for the authorized (having labels equal to 0) is mirrored around 0 and fit to a Gaussian distribution having mean 0. Then, we calculate the standard deviation σ of these samples and set the decision threshold to $\gamma = \min(0.5, 3\sigma)$. As for obtaining the ROC curve, we scan the value of γ from 0 to 1.

2) One Vs All (OvA): To obtain the ROC curve, we scan a single threshold γ from 0 to 1 such that $\boldsymbol{\gamma} = \gamma\mathbf{1}$. To calculate the accuracy, we use multiple thresholds designed according to the same method of Gaussian fitting used in Disc.

References

[1] P. L. Yu, J. S. Baras, and B. M. Sadler, "Physical-Layer Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 38–51, Mar. 2008.

[2] W. Wang, Z. Sun, S. Piao, B. Zhu, and K. Ren, "Wireless Physical-Layer Identification: Modeling and Validation," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 2091–2106, Sept. 2016.

[3] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu, and Y. Yan, "Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme," *IEEE Internet of Things Journal*, vol. 6, pp. 349–360, Feb. 2019.

[4] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Using the physical layer for wireless authentication in time-variant channels," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 2571–2579, July 2008.

[5] Q. Xu, R. Zheng, W. Saad, and Z. Han, "Device Fingerprinting in Wireless Networks: Challenges and Opportunities," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 94–104, Firstquarter 2016.

[6] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method," in *2011 Proceedings IEEE INFOCOM*, pp. 1404–1412, Apr. 2011.

[7] N. Gulati, R. Greenstadt, K. R. Dandekar, and J. M. Walsh, "GMM Based Semi-Supervised Learning for Channel-Based Authentication Scheme," in *2013 IEEE 78th Vehicular Technology Conference (VTC Fall)*, pp. 1–6, Sept. 2013.

[8] S. U. Rehman, K. Sowerby, and C. Coghill, "Analysis of receiver front end on the performance of RF fingerprinting," in *2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC)*, pp. 2494–2499, Sept. 2012.

[9] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep Learning Convolutional Neural Networks for Radio Identification," *IEEE Communications Magazine*, vol. 56, pp. 146–152, Sept. 2018.

[10] J. Yu, A. Hu, G. Li, and L. Peng, "A Robust RF Fingerprinting Approach Using Multi-Sampling Convolutional Neural Network," *IEEE Internet of Things Journal*, pp. 1–1, 2019.

[11] S. Gopalakrishnan, M. Cekic, and U. Madhow, "Robust Wireless Fingerprinting via Complex-Valued Neural Networks," *arXiv:1905.09388 [cs, eess, stat]*, May 2019.

[12] G. Baldini, C. Gentile, R. Giuliani, and G. Steri, "Comparison of techniques for radiometric identification based on deep convolutional neural networks," *Electronics Letters*, vol. 55, no. 2, pp. 90–92, 2019.

[13] I. Agadakos, N. Agadakos, J. Polakis, and M. R. Amer, "Deep Complex Networks for Protocol-Agnostic Radio Frequency Device Fingerprinting in the Wild," *arXiv:1909.08703 [cs, eess]*, Sept. 2019.

[14] Q. Wu, C. Feres, D. Kuzmenko, D. Zhi, Z. Yu, X. Liu, and X. 'Leo' Liu, "Deep learning based RF fingerprinting for device identification and wireless security," *Electronics Letters*, vol. 54, pp. 1405–1407, Nov. 2018.

[15] K. Merchant, S. Revay, G. Stantchev, and B. Nousain, "Deep Learning for RF Device Fingerprinting in Cognitive Communication Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, pp. 160–167, Feb. 2018.

[16] S. S. Hanna and D. Cabric, "Deep Learning Based Transmitter Identification using Power Amplifier Nonlinearity," in *2019 International Conference on Computing, Networking and Communications (ICNC)*, pp. 674–680, Feb. 2019.

[17] K. Youssef, L.-S. Bouchard, K. Z. Haigh, H. Krovi, J. Silovsky, and C. P. V. Valk, "Machine Learning Approach to RF Transmitter Identification," *arXiv:1711.01559 [cs, eess, stat]*, Nov. 2017.

[18] C. Geng, S.-j. Huang, and S. Chen, "Recent Advances in Open Set Recognition: A Survey," *arXiv:1811.08581 [cs, stat]*, July 2019.

[19] L. Shu, H. Xu, and B. Liu, "DOC: Deep Open Classification of Text Documents," *arXiv:1709.08716 [cs]*, Sept. 2017.

[20] D. Raychaudhuri, I. Seskar, M. Ott, S. Ganu, K. Ramachandran, H. Kremo, R. Siracusa, H. Liu, and M. Singh, "Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols," in *Wireless Communications and Networking Conference, 2005 IEEE*, vol. 3, pp. 1664–1669, IEEE, 2005.

[21] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over the Air Deep Learning Based Radio Signal Classification," *arXiv:1712.04578 [cs, eess]*, Dec. 2017.

[22] Y. Sun, A. K. C. Wong, and M. S. Kamel, "CLASSIFICATION OF IMBALANCED DATA: A REVIEW," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, pp. 687–719, June 2009.