INL/CON-24-81957-Revision-0

Contrasting Time-Frequency Representations for Unknown Waveform Detection

June 2025

Changing the World's Energy Future

Xue Wei, Dola Saha, Anna Quach, Daniel E. Wells

Idaho National Laboratory

INL/CON-24-81957-Revision-0

Contrasting Time-Frequency Representations for Unknown Waveform Detection

Xue Wei, Dola Saha, Anna Quach, Daniel E. Wells

June 2025

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Contrasting Time-Frequency Representations for Unknown Waveform Detection

Xue Wei[†], Anna Quach[*], Daniel Wells[*], and Dola Saha[†]
[†]Department of Electrical and Computer Engineering, University at Albany, SUNY
[*]Computational Data Science, Idaho National Laboratory
Email: [†]xwei4@albany.edu, [*]anna.quach@inl.gov, [*]daniel.wells@inl.gov, [†]dsaha@albany.edu

***Abstract*—Identifying unseen electromagnetic waveforms is critical for many applications, like interference management, electronic warfare and spectrum management. Traditionally this is done using statistical methods for anomaly detection, which has evolved to deep learning models for identifying the unseen data, formally termed as open set recognition. Some prior methods use a generative model to emulate open set data, which face challenges in generating synthetic samples for open set while simultaneously selecting an optimal discriminator for accurate classification. To alleviate this issue, we propose a discriminative model that effectively combines time and frequency domain features of communication signals for accurate predictions. We further introduce a cosine similarity loss that makes the domain specific features unique to enhance the prediction rate. Additionally, our model avoids generic feature vectors by extracting class-specific features during training, resulting in improved class representation. The experiment results show that this combined feature approach with cosine loss outperforms single-domain models and improves accuracy by 10% over models without cosine loss.**

***Index Terms*—Signal Intelligence (SIGINT), Open Set Recognition (OSR), Unknown Signal Detection**

I. Introduction

Accurate identification of novel electromagnetic signals has far reaching implications in different applications. It is essential to identify new waveforms, specially the war-reserve ones in an electronic warfare to enhance intelligence, surveillance, and reconnaissance (ISR) capabilities. It can also have real impact in identifying new electromagnetic signals and flagging them as radio frequency interference (RFI) for passive sciences, like radio astronomy and remote sensing. Current research in this field is limited to using signal processing based methods [1] to detect anomalies in the known system. However, the performance of these techniques suffer in accuracy of detection. Furthermore, often human intervention is required to set thresholds for classifying new waveforms.

Deep learning based methods have garnered attention for classification [2], [3] tasks. These models perform classification with high precision when the test dataset has the same distribution as the training dataset, more precisely known as the *close* set. Although they have high precision on the close set, they have no means to categorize samples that do not belong to the known classes, typically termed as the *open* set. Open set recognition (OSR) [4], [5] delves into this problem to accurately separate open from close datasets. This is generally achieved in two ways: generative or discriminative. Generative models often use Generative Adversarial Networks (GANs) [6], where the generator is supposed to generate out-of-order distribution of the close set while the discriminator learns to separate the known (true) and unknown (generated) data [7]–[9]. The success of this method lies in a) the generation of samples that represents a wide range of possible open set data and b) careful choice of the intermediate discriminator model when the generator has succeeded in representing the open set and the discriminator has been trained to be powerful enough to differentiate open vs close set data. Achieving both the objectives in the same training iteration is not guaranteed, which may result in suboptimal models. The discriminative approach [10], [11] aims to develop a highly accurate classifier model for the known set and subsequently creates an effective threshold to perform OSR. OpenMax [10] is the most prominent model in the discriminative approach, which simply extends the output of a neural network classifier to better handle unknown classes. One of the drawbacks of these techniques is that these models do not enhance the feature representation. Consequently, when candidates from the open set have a high resemblance in the extracted feature set with that of the close set data, it results in misclassification.

Most of the prior work in OSR is developed for image, video, or text data. Comparatively, there is limited prior work in the field of OSR handling electromagnetic signals. Some research exists in identifying unknown radio hardware for fingerprinting [12] and unknown modulation detection [13]. Research on identifying unknown electromagnetic waveforms is far less. One of the earlier works [14] attempts to take a discriminative approach using a simple threshold for accuracy in a discriminative model. This limits its effectiveness, particularly when the SNR is low. In other research [15], authors take a generative approach using a multi-task architecture providing both signal and modulation information for training. This may improve generalization, but is not quite effective in lower SNRs, where modulation identification is a harder task than detection itself. In our prior work [16], we embraced a generative approach through fusing multi-domain features of the same signal to detect unknown waveforms. However, since the model is developed using a data driven approach, it may generate similar features for each of the domains, even when it started with different domains of data. This is possible as the network may find the best feature set to represent the waveform. However, this defies the purpose of choosing different transforms of the same signal, where our hypothesis was that a sample from the open set may have similar features with the close set in one of the domains, but will not have similar feature sets in different domains. The model also relied on the knowledge of the start of the packet, which is difficult to estimate in low SNR and limits its applicability to packet oriented communication signals only. Furthermore, the model [16] also suffers from limitations common to a generative approach, as discussed above.

To overcome the challenges stated above, we propose a discriminative approach for detecting unknown waveforms. Communication signals are either modulated in the time domain or in the frequency domain, which makes the waveform features unique in that form of representation. This is further illustrated in § II, where the choice of the dataset that the model is trained on, affects the performance of the model. For some datasets, the time domain features dominant in providing high accuracy, while in other datasets, the frequency domain features perform better. Hence, it is essential to combine both time and frequency domain representation of the waveforms for OSR. We not only choose different domains of the waveforms, we also separate the domain specific features. Each domain feature undergoes different transformations while preserving enough information to be able to independently differentiate between the close and open set.

The contributions of this paper can be summarized as:

1) Feature Fusion: Our model integrates both time and frequency domain features to enhance performance while preserving their distinctiveness.
2) Contrasting Domain Specific Features: The model reduces similarity between the time and frequency domain features, which is essential for improving detection of unknown data.
3) Class Specific Feature Extraction: Instead of relying on the same feature vector for all classes, we create a feature set for each of the classes, which is maximized for each of the known classes/waveforms.
4) Wide Operational Range: We evaluate the classification performance on both close and open set data across a broad range of common communication waveforms and SNRs, where combining features with cosine similarity loss provides a performance improvement of 10% over the model trained without cosine similarity loss.

II. Problem Formulation

One of the major challenges of OSR using a data driven approach is the choice of data. During training, models only learn intricacies presented in the training data and cannot capture the features in unseen data. This is evident in Figure 1. We trained a multi-class classifier neural network model on two different datasets, $D1$ and $D2$ with time and frequency domain representation. This results in four unique models, two of them trained using time domain signals and two others using frequency domain signals. $D1$ consists of WLAN non-HT, LTE and Bluetooth LE, whereas $D2$ has Zigbee, DSSS and LTE. Time domain representation performs highly accurately with $D2$, but fails with $D1$. On the other hand, frequency domain representation performs much better with $D1$ than with $D2$. The performance is highly correlated with how the signal is modulated in each of the waveforms in the two datasets and how they manifest in time and frequency domains. If any one of the models (e.g. time domain model trained with $D2$) is chosen as the base model to extrapolate for OSR, it may result in misclassification for OSR, which will have data not seen during training. The problem is to a) create a model that captures both time and frequency domain features and b) curate a close dataset with a mix of known classes, where some may have prominent features in time domain and others may show unique frequency domain features.

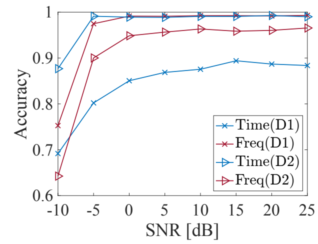


Fig. 1: Comparison of accuracy in time and frequency domains using two different datasets.

If Z denotes the features of waveforms X, where they belong to K number of classes, y, then the problem is to maximize detection probability P for both close and open set. We separate the detection probability into multi-class close and open set scenarios, represented as follows:

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{K} P(Z^{kk_i})P(y|Z^{kk_i}) + P(Z^{uu})P(y|Z^{uu}) \\ \text{subject to} \quad & 0 < P(Z) < 1, \text{and } 0 \leq P(y|Z) \leq 1, \text{and } K \in \mathbf{N}, \end{aligned} \tag{1}$$

where i is the index of each waveform class, kk and uu represents *known known* and *unknown unknown* classes [4] respectively and $P(y|Z)$ measures the probability of correctly classifying a label given feature Z. By improving $P(y|Z)$, we enhance the probability of accurate label prediction, which in turn contributes to higher classification accuracy.

III. System Design for Proposed Model

In this section, we introduce our method of solving the problem as formulated in Eq 1 in two main stages: a) Close Set Classification and b) Open Set Detection. The complete process is shown in Figure 2.

A. Close Set Classification

Our approach is based on the principle that achieving high classification accuracy on close set data is essential for effectively distinguishing between close and open set data. We introduce our domain knowledge for communication to extract features (Z) using both time and frequency domain representations. Based on the discussions in § II, we argue that both

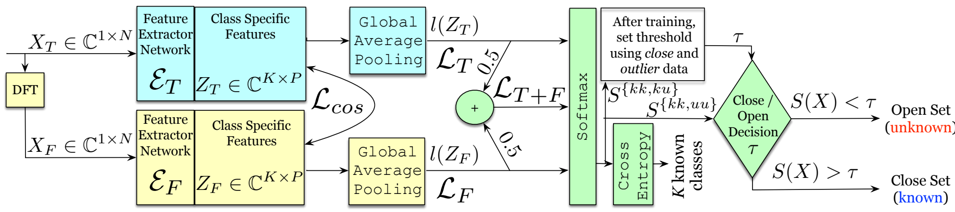


Fig. 2: The architecture of the proposed system developed by extracting both time and frequency domain features and contrasting them to improve performance of detection of unknown waveforms.

time and frequency domain representations are essential to accurately detect the communication waveforms. Combining the two complementary forms of signal representations, which are the primary modes of embedding bits to electromagnetic signals, we can capture essential factors of differentiating open and close set. We denote the time domain signal as $X_T \in \mathbb{C}^{1\times N}$, and its frequency domain representation can be obtained using the Discrete Fourier Transform (DFT):

$$X_F = DFT(X_T) = \sum_{n=0}^{N-1} X_T \cdot e^{-j2\pi n \frac{k}{N}}, \tag{2}$$

where n is the index of each time sample, N is the total number of samples in the signal, j is the imaginary component of the complex number, and k is the frequency bin index. Time domain features effectively capture temporal patterns and variations, while frequency domain features provide insights into the signal's spectral content and periodic characteristics. Hence, we define two feature extractor models, denoted as $\boldsymbol{\mathcal{E}}_T(\theta_T, X_T)$ for time domain and $\boldsymbol{\mathcal{E}}_F(\theta_F, X_F)$ for frequency domain features, respectively. θ_T and θ_F are model parameters. These two models are responsible for extracting class-specific features $Z_T \in \mathbb{C}^{K\times P}$ and $Z_F \in \mathbb{C}^{K\times P}$ independently, where K represents the number of close set classes and P denotes the feature length. To ensure that Z_T and Z_F represent different characteristics of the original signal and to maximize the use of the signal information, we minimize the similarity between the time and frequency domain features by reducing their cosine similarity, $\min(\cos(Z_T, Z_F))$. In this way, these two sub-classifiers are forced to make correct classifications at the same time based on different characteristics of the signal. Let y denote the correct label and I represent the mutual information. Then, $I(Z; y)$ captures the information relevant to classification. The relationship between the features Z_T and Z_F can be expressed as follows:

$$I(Z_T; y) \neq I(Z_F; y)\,; I(Z_T; y) > 0\,; I(Z_F; y) > 0 \tag{3}$$

The inequality emphasizes the distinct contributions of each feature domain to the mutual information with the class label y. Consequently, the combined feature representation $Z = (Z_T, Z_F)$ offers a more comprehensive depiction of the original signal X, resulting in enhanced classification performance. Specifically, the total mutual information between the combined features and y surpasses that of either feature alone, which can be expressed mathematically as follows:

$$I(Z; y) > I(Z_T; y), \quad I(Z; y) > I(Z_F; y). \tag{4}$$

The conditional entropy $H(y|Z)$ quantifies the uncertainty remaining about y after observing Z:

$$H(y|Z) < H(y|Z_T), \quad H(y|Z) < H(y|Z_F). \tag{5}$$

We can then conclude that:

$$H(y|Z) < \min(H(y|Z_T), H(y|Z_F)) \tag{6}$$

indicating that the combined features reduce the uncertainty in predicting labels, thereby enhancing classification accuracy.

We implement the feature extractor models, $\boldsymbol{\mathcal{E}}_T(\theta_T, X_T)$ and $\boldsymbol{\mathcal{E}}_F(\theta_F, X_F)$ as two neural networks, which are described in details in § IV. The extracted features, Z_T and Z_F, first undergo global average pooling to produce logits. The logits from both the time and frequency domains are then combined and passed through a softmax activation function to make the final classification outcome, as illustrated in Figure 2.

B. Open Set Detection

In this stage, we use the trained classifier and establish an appropriate threshold τ to extend its capability for OSR. To ensure methodological rigor and data integrity, it's crucial to differentiate between open set data used for threshold setting and data used for testing. Specifically, we use X^{ku}, referred to as "known unknown" open set data, for threshold setting, while X^{uu}, representing "unknown unknown" open set data, for which no prior information is available and are reserved for the testing phase only.

During the neural network training process for multi-class classification, each neuron in the output layer produces a probability for its corresponding class, denoted as $P(y = i|X)$ for class i. As we use `softmax` activation function in the output layer, the probability distribution over all classes satisfies:

$$\sum_{i=1}^{K} P(y = i|X) = 1, \tag{7}$$

The predicted class $\hat{y}$ is selected based on the maximum predicted probability:

$$\hat{y} = \arg\max_i P(y = i|X). \tag{8}$$

For open set data, since these samples are not represented in the training set, the classifier is likely to struggle with assigning a high probability to any specific class. Consequently, the maximum probability for open set data, is generally expected to be lower than that for close set data. Thus, for each data sample X, we define the score $S(X)$ as follows:

$$S(X) = \max_i \; l(Z_{T+F}), \tag{9}$$

where $l(Z_{T+F})$ is the combined logits by averaging the $l(Z_T)$ and $l(Z_F)$, with equal weights of 0.5, reflecting the equal importance of time and frequency domains in our analysis:

$$l(Z_{T+F}) = 0.5 \cdot l(Z_T) + 0.5 \cdot l(Z_F). \tag{10}$$

To establish a suitable threshold, the Youden's Index (J) is calculated and defined as:

$$J = \text{Sensitivity} + \text{Specificity} - 1, \tag{11}$$

where Sensitivity $= \frac{\text{True Positive}}{\text{True Positive+False Negative}}$ and Specificity $= \frac{\text{True Negative}}{\text{True Negative+False Positive}}$. The Youden index balances sensitivity and specificity and helps identify an optimal threshold by maximizing the difference between true positive (TP) and false positive (FP) rates. We set the threshold, τ, at the value that maximizes the Youden index, defined as

$$\tau = \arg\max_{\tau} J \tag{12}$$

This threshold minimizes misclassification between open set and closed set data, thereby improving the model's ability to distinguish between known and unknown samples.

In the testing phase, for each input sample, X, we decide whether it is classified as known or unknown according to the following equation:

$$\begin{cases} X \in \text{known}, X \in \arg\max_i l(Z_{T+F}) & \text{if } \mathrm{S}(X) \geq \tau, \\ X \in \text{unknown} & \text{if } \mathrm{S}(X) < \tau. \end{cases} \tag{13}$$

When evaluating open set problems, we can differentiate the conditional entropy into multi-class and open set scenarios, expressed by conditioning on each class as follows:

$$H(y|X) = \sum_{i=1}^{K} P(Z^{kk_i})H(y|Z^{kk_i}) + P(Z^{uu})H(y|Z^{uu}) \tag{14}$$

Here, $P(Z)$ is defined based on the characteristics of the input dataset. Additionally, $H(y|Z)$ in Equation 6 is minimized by leveraging combined features across classes in both the close set and open set. This broader feature representation enhances the model's ability to capture a comprehensive range of information from the original signal. As a result, the reductions in $H(y|Z^{kk})$ and $H(y|Z^{uu})$ correspond to increased probabilities of accurate label predictions, thereby improving $P(y|Z^{kk})$ and $P(y|Z^{uu})$ as presented in Equation 1. This approach effectively reduces classification uncertainty and improves overall accuracy.

IV. Neural Network Architecture

A. Model Structure

We design two feature extractor networks, $\boldsymbol{\mathcal{E}}_T(\theta_T, X_T)$ for time domain and $\boldsymbol{\mathcal{E}}_F(\theta_F, X_F)$ for frequency domain feature extraction. The models have been developed using a complex neural network [17] with identical layers, as shown in Table I. The structure comprises of three complex convolutional layers, each of them is followed by a normalization layer and a complex leaky ReLU function. Layer normalization is chosen here to stabilize and accelerate the training process, and it is also independent of batch size. It is worth noting that the features after the convolutional layer are generated per class, with one vector for each class. This design maps the original signal into a feature space representing K distinct classes, where the highest feature value corresponds to the true class. The extracted features are then processed by a global average pooling layer, which averages the logits $\{l(Z_T), l(Z_F)\}$ of the features $\{Z_T, Z_F\}$ and are used to make classification.

TABLE I: Neural Network Architecture.

Identical Layers in $\boldsymbol{\mathcal{E}}_T$ and $\boldsymbol{\mathcal{E}}_F$	**Dimension**	**Neg Slope**
Conv1d + LayerNorm + C_LReLU	[2,250]	0.3
Conv1d + LayerNorm + C_LReLU	[3,125]	0.3
Conv1d + LayerNorm + C_LReLU	[3,125]	0.3
AvgPool1d	[3,1]	**Batch Size**
Softmax	[3,1]	200

B. Loss Function

Our model is trained for close set classification under the condition that the time and frequency features extracted have distinct effects. It is essential to separate the features in an effort to find multiple representations that are individually capable of performing the classification accurately on the close set. Having multiple unique features of the same signal ensures that unseen data may have similarity with one of the features from known class, but not both of them. To make Z_T and Z_F unique that represent different aspects of X, we minimize their cosine similarity. When $\cos(\theta) = 0$, the vectors are orthogonal, $Z_T \perp Z_F$, implying they are uncorrelated and capture distinct features. The cosine similarity loss function between Z_T and Z_F can be expressed as:

$$\mathcal{L}_{cos} = \sum_{i=1}^{K} \frac{Z_T^i \cdot Z_F^i}{||Z_T^i|| \cdot ||Z_F^i||} \tag{15}$$

The objective of the classification model is to ensure that the predicted label $\hat{y}$ closely aligns with the true label y, i.e., the model aims to achieve $\hat{y} = y$ for all X. To this end, we desire that the distribution of predicted labels $P(\hat{y}|X)$ matches the distribution of the true labels $P(y|X)$. The cross-entropy, $H(y, \hat{y})$, is employed to measure the dissimilarity between these two distributions. For a multi-class classification problem with K classes, this is formalized as:

$$H(y, \hat{y}) = -\sum_{i=1}^{K} P(y = i|X) \log(P(\hat{y} = i|X)), \tag{16}$$

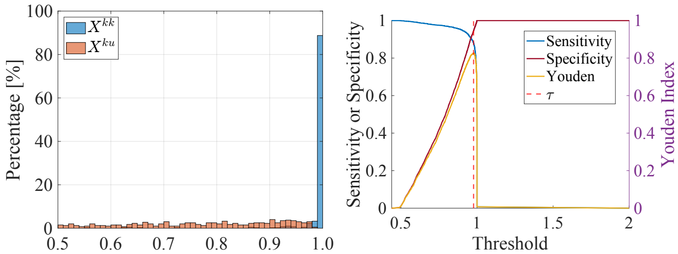
where $P(y = i|X)$ is the true probability of class i for signal X and $P(\hat{y} = i|X)$ is the predicted probability of class i for signal X. To ensure our model effectively leverages both the time domain and frequency domain features, we apply loss functions to time domain features, frequency domain features, and combined features respectively. The softmax function is applied before computing the cross-entropy loss in our model. The loss function $\mathcal{L}_{\mathcal{T}}$ can be expressed as:

$$\mathcal{L}_{\mathcal{T}} = -\sum_{i=1}^{K} P(y = i|X) \log\left(\frac{e^{l(Z_{T_i})}}{\sum_{j=1}^{K} e^{l(Z_{T_j})}}\right) \tag{17}$$

Similarly, the loss functions $\mathcal{L}_{\mathcal{F}}$ and $\mathcal{L}_{\mathcal{T}+\mathcal{F}}$ can be derived by replacing Z_T with Z_F and $Z_{(T+F)}$ respectively. The overall

TABLE II: Training Parameters.

Parameter	Value	Parameter	Value
Learning rate	0.0004	Epochs	500
Channel SNR	-10 $\sim$ 25 dB	Platform	Pytorch
Open (X^{ku}) for τ	Zigbee	Open (X^{uu}) / Test	DSSS
Close set (X^{kk})	WLAN non-HT, Bluetooth LE, LTE		
Data position	Random position within the signal		



(a) Score, S(X), of close and open set at 25 dB SNR.

(b) Youden's Index to set the threshold, τ.

Fig. 3: Evaluation process for setting a Threshold.

loss function can be decomposed into combined and domain-wise cross-entropy loss terms and the diversity term, which can be formulated as:

$$\mathcal{L}_{Total} = \mathcal{L}_{cos} + \mathcal{L}_{T} + \mathcal{L}_{F} + \mathcal{L}_{T+F} \tag{18}$$

V. Experiments

Data Generation: We utilize MATLAB generated wireless communication signals to create five types of waveforms: WLAN non-high throughput (non-HT), Bluetooth Low Energy (LE), Long-Term Evolution (LTE), Zigbee, and Direct Sequence Spread Spectrum (DSSS). WLAN non-HT and LTE use Orthogonal Frequency Division Multiplexing (OFDM), with WLAN non-HT being packet-based and LTE frame-based. Bluetooth LE employs Frequency-Hopping Spread Spectrum (FHSS), while Zigbee uses O-QPSK modulation. DSSS technology is applied in 802.11b. Each dataset undergoes additive white Gaussian noise (AWGN) across various SNRs, with 10,000 I/Q samples per SNR level.

Experiment Setup: The experiment setup, including the training parameters, are shown in Table II. WLAN non-HT, Bluetooth LE, and LTE are used in the training phase as the closed set datasets. The Zigbee dataset is included as the known open set data for threshold setting, τ, while 802.11b (DSSS) is designated as the unknown open set dataset for performance testing. To optimize the amount of data we have, we use a ratio of 8:1:1 for the close set data for training, threshold setting (τ), and testing, respectively. A normalization process is applied before passing the data to the model to ensure consistent scale and improved convergence speed. The neural network models are designed and optimized using PyTorch.The networks are trained for 500 epochs using the Adam optimizer with a learning rate of 0.0004.

VI. Results

A. Threshold setting

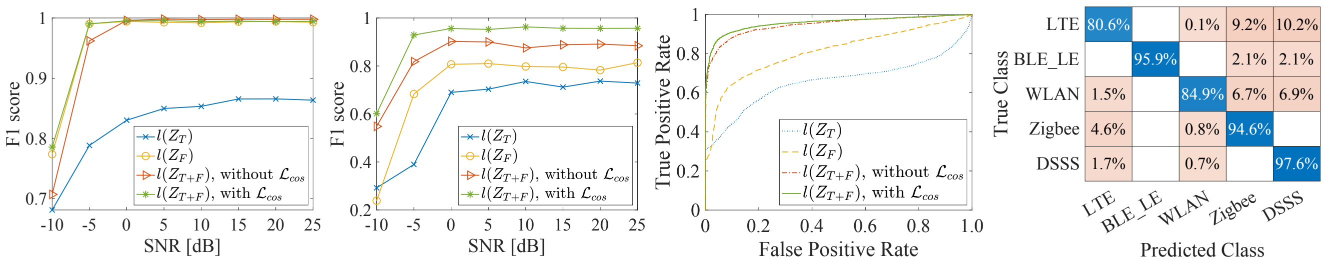
The choice of the threshold, τ, significantly impacts the performance of OSR. For applications where minimizing open set misclassification is critical, a higher threshold can reduce open set samples being incorrectly classified as close set, though it may also lead to some close set samples being misclassified. Conversely, a lower threshold helps maintain higher accuracy for close set data but risks classifying more open set samples as close set. To understand the behavior of the score $S(X)$, on which the threshold will be calculated, we inspect it for both X^{kk} and X^{ku} at a reasonably high SNR, 25 dB, as shown in Figure 3a. A clear separation is evident between the score distributions of X^{kk} and X^{ku}. Since the input is selected randomly from any position within the signal, $S(X^{ku})$ is not concentrated at any specific point. By choosing an optimal threshold, we can effectively distinguish open set samples from X^{uu}. Figure 3b illustrates sensitivity, specificity, and Youden's index across different thresholds. Here, we select the threshold value ($\tau = 0.9774$) corresponding to the maximum Youden's index (0.8266), as it maximizes the difference between TPR and FPR, providing an optimal balance for distinguishing between known and unknown samples. We use this value of τ in the rest of the paper for performance evaluation of OSR.

B. Close Set Classification

Figure 4a illustrates the close set classification performance (F1 score) of our proposed model, labeled as '$l(Z_{(T+F)})$, with $\mathcal{L}_{cos}$', which is compared with three other *baseline cases*, labeled as: a) '$l(Z_T)$' and b) '$l(Z_F)$', where our final trained model is used with decisions made only with time and only with frequency domain features, respectively, and c) '$l(Z_{(T+F)})$, without $\mathcal{L}_{cos}$', is trained with the same model with combined features, excluding the cosine similarity loss. The frequency domain and combined features yield similar and superior F1 scores compared to the model with time domain features only. This can be attributed to the fact that the data may come from any position within a packet, not necessarily containing the preamble, making the time domain classification more challenging, especially as SNR decreases. At lower SNRs, frequency domain features alone perform well for classification of close set; however, time domain features play a crucial role in enhancing OSR performance, which will be discussed in the following section.

C. Open Set Detection

Next, we measure the predictive performance of the models on both close and open set data. Figure 4b shows the F1 score of our model under different SNR levels and distinguishing open set based on different kinds of features. Using combined features with the cosine loss demonstrates the best performance, as the cosine loss effectively preserves feature distinctiveness, enhancing the model's predictive ability. Although combining features alone improves prediction, the absence of cosine loss reduces the ability to fully leverage the feature variation, resulting in lower performance compared to when cosine loss is used. Although frequency domain feature can achieve high close set classification accuracy as evident in Figure 4a, it is not enough for open set detection. Therefore, it is necessary to combine time and frequency features to obtain



(a) Classifying close set. (b) Classifying open/close set. (c) OSR AUROC curve; all SNRs. (d) OSR fused model with $\mathcal{L}_{cos}$.

Fig. 4: Multi-class classification for varying SNRs and OSR performance for all SNRs.

higher performance for OSR. Overall, combining features with cosine loss improves the F1 score by 10%, 20%, and 30% compared to combining features without cosine loss, frequency features only, and time features only, respectively. Additionally, as SNR increases, the F1 score improves across all feature types since there is less noise in the signal.

We also compare the ROC curve achieved in our proposed method across all SNRs and compared with three baselines, as shown in Figure 4c. The curves for combined features with and without cosine loss are close, though the inclusion of cosine loss slightly increases the area under the curve. Overall, ROC curves for combined features are consistently higher than those for individual features alone, indicating superior performance in distinguishing classes.

Finally, we examine the classification performance for each class individually. Figure 4d shows the confusion matrix for close and open set predictions across all SNR levels using combined features with cosine loss. Since we set a relatively high threshold, most open set samples are correctly identified, demonstrating good performance in rejecting open set samples. However, some close set samples are misclassified as open, specially at very low SNRs as seen in the confusion matrix.

VII. Conclusion

Detection of unknown signals and new waveforms is a critical problem in electronic warfare. We proposed a novel approach to perform multi-class classification on the known signals of interests, while simultaneously performing binary classification for differentiating the known from the unknown signals. We combine time and frequency domain features and make them unique for a robust detection of unknown signals. Future work will include the usage of more diverse signals for both the open and close set, and signals captured over-the-air.

Acknowledgment

This research is supported in part by Idaho National Laboratory (INL) and the United States Department of Energy (DoE) under Lab Directed Research and Development (LDRD) grant 24A1081-049FP and 296512. We thank Nicholas Kaminski from INL for his consultation throughout the project.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, jul 2009.

[2] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[4] C. Geng, S. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, pp. 3614–3631, oct 2021.

[5] A. Mahdavi and M. Carvalho, "A survey on open set recognition," in *2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, pp. 37–44, 2021.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27* (Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, eds.), pp. 2672–2680, Curran Associates, Inc., 2014.

[7] S. D. Zongyuan Ge and R. Garnavi, "Generative openmax for multi-class open set classification," in *Proceedings of the British Machine Vision Conference (BMVC)* (G. B. Tae-Kyun Kim, Stefanos Zafeiriou and K. Mikolajczyk, eds.), pp. 42.1–42.12, BMVA Press, September 2017.

[8] S. Kong and D. Ramanan, "Opengan: Open-set recognition via open data generation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–10, 2022.

[9] G. Chen, P. Peng, X. Wang, and Y. Tian, "Adversarial reciprocal points learning for open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 8065–8081, 2022.

[10] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1757–1772, 2013.

[11] A. Rozsa, M. Günther, and T. E. Boult, "Adversarial robustness: Softmax versus openmax," 2017.

[12] P. Yin, L. Peng, G. Shen, J. Zhang, M. Liu, H. Fu, A. Hu, and X. Wang, "Multi-channel cnn-based open-set rf fingerprint identification for lte devices," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 5, pp. 1788–1800, 2024.

[13] M. Zhang, P. Tang, G. Wei, X. Ni, G. Ding, and H. Wang, "Open set domain adaptation for automatic modulation classification in dynamic communication environments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 10, no. 3, pp. 852–865, 2024.

[14] S. R. Shebert, A. F. Martone, and R. M. Buehrer, "Open set wireless standard classification using convolutional neural networks," in *MILCOM 2021 - 2021 IEEE Military Communications Conference (MILCOM)*, pp. 757–762, 2021.

[15] J. Gong, X. Qin, and X. Xu, "Multi-task based deep learning approach for open-set wireless signal identification in ism band," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 1, pp. 121–135, 2022.

[16] X. Wei, D. Saha, and A. Quach, "Exploiting multi-domain features for detection of unclassified electromagnetic signals," in *MILCOM 2024 - 2024 IEEE Military Communications Conference (MILCOM)*, 2024.

[17] C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," *CoRR*, vol. abs/1705.09792, 2017.