Expert Fusion with Meta-Adaptation for Signal Technology Recognition

Ife Olalekan Ebo[1], Idowu Ajayi[1], Lina Mroueh[1], and Youmni Ziade[2]
[1]*Lab. d'Informatique, Signal et Image, Télécommunications et Élect. (LISITE), Isep, 92130 Issy-les-Moulineaux, France.*
[2]*L'Agence Nationale des Fréquences (ANFR), 94704 Maisons-Alfort, France.*
Corresponding author: ife-olalekan.ebo@isep.fr

***Abstract*—Low Power Wide Area Networks (LPWANs) enable cost-effective, low power consumption, and long-range IoT applications but often operate under low Signal-to-Noise Ratio (SNR) conditions. Conventional deep learning models for signal recognition struggle to generalize in such dynamic environments. We propose a lightweight framework combining ensemble learning, a Mixture of Experts (MoE) with uncertainty-aware soft gating, and meta-adaptation using the Almost No Inner Loop (ANIL) method. Expert models trained at distinct low-SNR levels share frozen base weights, while only the classifier head is adapted during few-shot learning for rapid specialization. The uncertainty-aware gating mechanism produces sparse, temperature-controlled expert weights, enhancing robustness and decision reliability. Experiments demonstrate improved accuracy and generalization across both seen and unseen SNR levels, highlighting the framework's effectiveness for real-world wireless signal technology recognition in spectrum monitoring systems.**

***Index Terms*—Deep learning, LPWAN, meta-learning, mixture of experts, spectrum monitoring, signal recognition.**

I. Introduction and Motivations

The widespread adoption of Internet of Things (IoT) devices, sensors, and technologies has profoundly impacted various sectors, becoming integral to modern daily life. The global number of connected devices is growing at a remarkable pace, with a 13% increase recorded in 2024, bringing the total to 18.8 billion. This upward trend is expected to continue, with estimates suggesting that by 2030, around 40 billion IoT devices will be in operation worldwide [1]. While short-range technologies like Bluetooth, Bluetooth Low Energy and WiFi are widely used, Low Power Wide Area Networks (LPWAN) are becoming increasingly important for IoT applications that require long-range connectivity [2]. Operating in the sub-GHz unlicensed Industrial, Scientific, and Medical (ISM) bands, LPWAN offers numerous advantages, including low power consumption, extended range, free access to radio spectrum, cost-effectiveness, long battery life, and scalability [3].

LPWANs play a critical role in enabling IoT applications such as in smart health, smart cities, smart agriculture, smart industry, smart logistics, etc [4]. These technologies coexist and operate under strict power and bandwidth constraints, making them highly susceptible to interference, fading, noise, most especially at low Signal-to-Noise-Ratio (SNR) environments. These issues make it challenging to identify them in the wireless spectrum. To address these challenges, previous research has explored various approaches, including the use of machine learning classifiers [5]. Moreover, deep learning models have shown strong success using Convolutional Neural Networks (CNN) [6], Recurrent Neural Networks (RNN) [7], Transformer-based models [8], and a combination of any of these models [9]. However, traditional deep learning models often struggle and fail with adaptability, and generalization in dynamic environments where SNR levels vary significantly, particularly in low-SNR levels environment. These limitations make them less suitable for certain real-world applications [10].

The Mixture of Experts (MoE) is an ensemble learning approach designed to improve predictive performance by training specialized models, or experts, on distinct subsets of data. Each expert focuses on a specific subtask where it performs optimally, making MoE particularly effective for complex predictive modeling problems [11]. This technique has demonstrated its effectiveness in computer vision [12], deepfake speech detection [13], Large Language Models (LLMs) and its scaling [14], [15], and recommender systems [16].

Recent works in the State-of-the-Art (SOTA) [17]–[22] have explored the use of MoE framework alongisde transformer-based methods, residual neural networks, and CNN, for automatic modulation classification, technology recognition and anomaly detection. However, these models have been trained using datasets at some specific SNRs during model training and feature extraction. Meanwhile, the complexity of wireless environment in the real-world scenarios varies over time. This implies that the models may under-perform when faced with conditions not seen during training. In other words, they lack the ability to adapt, on-the-fly, to unseen environments. This raises the need for models capable of generalizing to previously unseen situations in the model training, catering for uncertainty with few samples.

To this end, we propose an uncertainty-aware expert fusion with meta-adaptation based on Almost No Inner Loop (ANIL)-based meta-learning framework. This approach enables few-shot adaptation for robust classification under dynamic SNR conditions without full retraining. It leverages frozen pretrained experts proposed in our work in [23], referred to as Channel Attention-based Denoising Autoencoder U-Net and Classifier (UNA-DAEC). Each expert, which makes the backbone, is trained at a unique SNR, and optimized through uncertainty-aware sparse gating and meta-learning. The resulting architecture is SNR-agnostic, delivers stable performance across diverse SNRs, and is suited for deployment in dynamic real-world spectrum monitoring systems.

This work was funded by LISITE, Isep and L'Agence Nationale des Fréquences (ANFR).

979-8-3503-9281-4/25/$31.00 ©2025 IEEE

The key contributions of this work are therefore as follows:

- We introduce an SNR-agnostic architecture for LPWAN signal technology recognition using a MoE model. Each expert has frozen, pre-trained weights, enabling generalization across varying signal types and noise conditions.
- We adopt an ANIL meta-learning strategy, adapting only the final classification layer through 5-shot learning. This allows fast adaptation to unseen SNR levels with minimal supervision and lower computational cost than full fine-tuning.
- We evaluate our approach on a real-world dataset from [24], showing strong performance under dynamic SNR conditions.

The remainder of this paper is structured as follows: Section II-A provides a background of the paper, and Section II-B provides an overview of the UNA-DAEC model. In Section III, we present our proposed model architecture. Section IV describes the analysis of the methodology used. Section V outlines the results and discussions. Section VI concludes the paper with insights into future research directions.

II. BACKGROUND

A. Background Description

In this section, we examine a scenario in which a spectrum monitor is deployed to collect data from IoT sensors operating across various LPWAN technologies coexisting within the same environment. In such real-world settings, multiple LPWAN signals coexists, along with background noise, resulting in composite received signals. The discrete-time received signal, $r[n]$, is expressed as:

$$r[n] = \sum_{i=1}^{N} s_i[n] + \eta[n], \quad \eta[n] \sim \mathcal{N}(0, \sigma_\eta^2), \tag{1}$$

where n is the discrete-time index, N is the number of active LPWAN transmitters, $s_i[n]$ is the transmitted signal from the i-th source, and $\eta[n]$ is i.i.d. additive white Gaussian noise (AWGN). To analyze $r[n]$, we compute its time-frequency representation (i.e. spectrogram) using the Short-Time Fourier Transform (STFT):

$$X[\mathrm{k}, m] = \sum_{n=0}^{M-1} r[n]\, h[n - mH]\, e^{-j\frac{2\pi \mathrm{k} n}{M}}, \tag{2}$$

where k is the frequency bin index ($\mathrm{k} \in \{0, 1, \ldots, M-1\}$), m is the time frame index, $h[n]$ is a Hamming window of length M, and H is the hop size (samples between consecutive windows). The spectrogram $\mathbf{S}[\mathrm{k}, m]$, representing signal energy, is derived in (3). It captures the signal's time-frequency energy distribution computed over all time and frequency bins.

$$\mathbf{S}[\mathrm{k}, m] = |\mathbf{X}[\mathrm{k}, m]|^2 . \tag{3}$$

The resulting 2D spectrogram matrix, $\mathbf{S}$, is used as the input to our proposed model, f_U, parameterized by U, to predict the signal class label, $\hat{\mathrm{y}}$ (e.g., LoRa, Sigfox, etc), as shown in (4).

$$\hat{\mathrm{y}} = f_U(\mathbf{S}[\mathrm{k}, m]). \tag{4}$$

B. Overview of UNA-DAEC Model

In our earlier work [23], we presented the UNA-DAEC model, a deep learning architecture designed to effectively denoise and classify LPWAN signals under challenging low-SNR conditions. The model's core objective is to enhance signal clarity in highly noisy environments, thereby achieving optimally accurate classification with a single forward and backward pass. Structurally, the model consists of a U-Net-based denoising autoencoder (DAE) on one side and a classifier block on the other. Within the U-Net DAE, channel attention mechanisms are integrated after each convolutional and deconvolutional layer, with the exception of the final deconvolution layer. Fig. 1 shows the simplified architecture with the three main components.

First, the channel attention block compresses global spatial information into a channel descriptor, effectively summarizing spatial features across the input feature map. Average pooling is used to reduce spatial dimensions and integrated into both the encoder and decoder sections of the U-Net architecture.

The U-Net architecture block, incorporates a convolutional DAE to transform noisy inputs into cleaner, reconstructed outputs through a structured encoding-decoding process.

Finally, the classifier block employs CNN to perform signal classification. By combining reconstruction and classification loss within a single optimization step, UNA-DAEC achieves improved performance compared to SOTA deep learning approaches. Moreover, leveraging time-frequency representations, such as spectrograms, proved to be more effective compared to other approaches. A more detailed explanation and its experimental results can be found in [23].

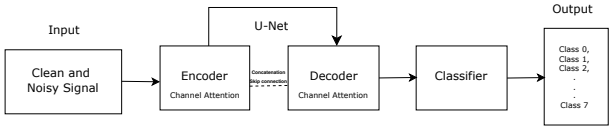

Fig. 1. A simplified architecture of the UNA-DAEC model, functioning as the expert.

III. Proposed Model

To address the earlier described limitations of models underperforming in unfamiliar environments, our proposed model for signal technology recognition in a varying SNR environments combines an ANIL meta-learning framework with a sparse MoE design. The MoE framework is a variant of the sparse MoE [19], integrating soft-gating with top-k expert selection. The benefit of this approach is that it balances efficiency (activating only a few experts), performance (allowing multiple experts to contribute) and scalability (allows flexibility in choosing the number of experts, k, balancing accuracy and computation). The proposed model leverages frozen pre-trained UNA-DAEC models, each specialized for a different SNR level, as the expert backbone. To enable generalization under previously unseen SNR conditions, we use 5-shot learning of the ANIL technique. ANIL is a simplified model agnostic meta-learning with lower computational overhead, and only requires updates on the final classification layer during adaptation [25]. The process

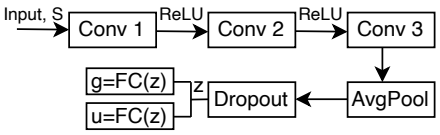

Fig. 2. Uncertainty-Aware Gating Architecture.

consists of UNA-DAEC as a pre-trained backbone model, as described in Section II-B, an uncertainty-aware gating network as shown in Fig. 2, and an expert fusion network based on meta-learning, described in Section III-C. The input to the model is $\mathbf{S} \in \mathbb{R}^{CH \times H \times W}$, a 1-channel spectrogram of size $H \times W$, where CH is the channel size, H is the height of 128 and W is width of 128.

A. Uncertainty-Aware Gating Network

When an input is received, the network dynamically selects among expert models by computing soft expert selection weights that are modulated by uncertainty. In other words, it improves expert selection under noisy or ambiguous conditions by assigning lower confidence inputs to more reliable or diverse experts. This enables contributions among experts. The gating network is composed of three convolutional layers (kernel size 3, stride 2, padding 1), followed by adaptive average pooling and then split into two distinct fully connected branches (i.e. gating, $\mathbf{g}$ (5) and uncertainty, $\mathbf{u}$ (6)) expressed as shown in Fig 2.

$$\mathbf{g} = \mathbf{W}_g \mathbf{z} + \mathbf{b}_g, \tag{5}$$

$$\mathbf{u} = \text{clamp}_{[0,u_{\max}]}(\mathbf{W}_u \mathbf{z} + \mathbf{b}_u), \tag{6}$$

where $\mathbf{z}$ is output from the average pooling, which serves as input to $\mathbf{g}$ and $\mathbf{u}$, $\mathbf{g} \in \mathbb{R}^N$ is the output gating logits for N experts, $\mathbf{u} \in \mathbb{R}^N$ is the uncertainty estimates per expert, $\mathbf{W}_{g/u}$ is the weight matrix for the gating/uncertainty, $\mathbf{b}_{g/u}$ is the bias term associated with the gating/uncertainty and $u_{\max}$ is the maximum uncertainty clamped at 5.

We adjust the sharpness of the softmax distribution applied on the gating weights based on the predicted uncertainty as in (7) and (8), and prevent expert collapse and exploding gradients by applying adaptive temperature scaling and weight clamping (minimum of 10^{-4}) as in (9).

$$\tau_i = T_0 + u_i. \tag{7}$$

$$\text{w}_i = \frac{\exp(g_i/\tau_i)}{\sum_{j=1}^{N} \exp(g_j/\tau_j)}. \tag{8}$$

$$\text{w}_i \leftarrow \text{clamp}_{[\text{w}_{\min}]}(\text{w}_i), \tag{9}$$

where τ_i is the temperature for expert i, and T_0 is fixed base temperature at 1.

Furthermore, we promote sparsity in expert usage by retaining only the top-k scoring experts per input, for more efficient computation and avoiding diffuse expert assignments. Given the expert weight vector, $\mathbf{w} \in \mathbb{R}^N$, for a batch size, B, the indices of the top-k, $\mathcal{I}_{\text{topK}}$ are identified for each input as (10) and a binary mask is used to show only top-k indices are active as (11). The mask $\mathbf{m}$, is applied to the original weights, $\mathbf{w}$, which results to $\tilde{\mathbf{w}}$ as (12) and the sparse weights, $\mathbf{w}^{\text{sparse}}$, are normalized to preserve proper convex combination as (13).

$$\mathcal{I}_{\text{topK}} = \text{TopK}(\mathbf{w}, k). \tag{10}$$

$$m_{b,i} = \begin{cases} 1 & \text{if } i \in \mathcal{I}_{topK}^{(B)} \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

$$\tilde{\mathbf{w}} = \mathbf{w} \odot \mathbf{m}. \tag{12}$$

$$\mathbf{w}^{\text{sparse}} = \frac{\tilde{\mathbf{w}}}{\sum_{i=1}^{N} \tilde{\text{w}}_i + \epsilon}, \tag{13}$$

where $\odot$ is Hadamard product (i.e. element-wise multiplication) and ϵ is a small constant $1e^{-8}$, added for numerical stability. This sparse gating strategy encourages the model to specialize by routing each input through only the most relevant subset of experts, which improves both computational efficiency and robustness in varying SNR conditions.

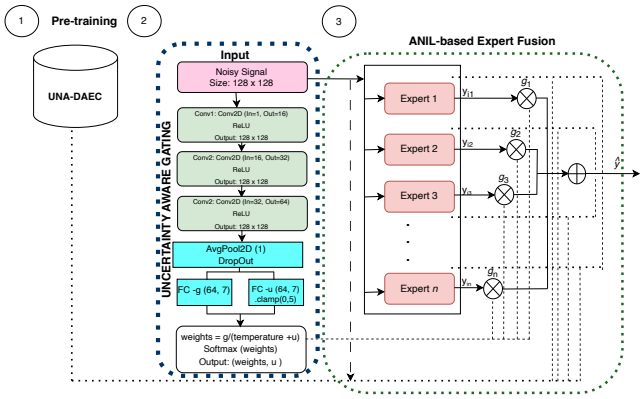

Fig. 3. Proposed Architecture.

B. Expert Network

These specialized neural network layers, known as experts, are designed to handle specific sub-tasks effectively. Each expert processes the same input but applies its learned specialization to extract relevant features, allowing the model to leverage diverse knowledge for improved performance. In this study, each expert is a separate instance of the UNA-DAEC model, trained on a unique subset of the dataset corresponding to a specific low-SNR level. The SNR values range from 0 dB to -20 dB, covering seven distinct levels: 0 dB, -3 dB, -6 dB, -10 dB, -13 dB, -16 dB, and -20 dB. Each selected expert $f_i(\mathbf{S})$ generates a prediction as shown in (14):

$$\tilde{\text{y}}_i = f_i(\mathbf{S}), \quad \forall i \in \mathcal{I}_{\text{topK}}, \tag{14}$$

where $\tilde{\text{y}}_i$ is the classification output of the i-th selected expert.

C. Expert Fusion based Meta-learning

The training process of the meta-learner incorporates the gating mechanism, a set of experts, and a top-k selection strategy. The uncertainty-aware gating module dynamically selects an optimal combination of experts based on the input signal. Each expert consists of a frozen denoising function trained at a specific SNR level, which processes the input without any further parameter updates. The denoised output is

then forwarded to a classifier. During adaptation, only the final classification layer (i.e., the classifier head) is updated, allowing the model to transform the denoised signal for optimally accurate recognition with minimal computational overhead.

All the denoised outputs, $\tilde{\mathbf{x}}_i$, and classification logits, $\mathbf{z}_i$, are stacked. The expert outputs are combined using weighted summation. The model returns the final denoised signal, $\mathbf{x}^{\text{final}}$, and the aggregated classification logits, $\mathbf{z}^{\text{final}}$, as:

$$\mathbf{x}^{\text{final}} = \sum_{i=1}^{N} \mathrm{w}_i \cdot \tilde{\mathbf{x}}_i = \text{sum}(\mathbf{w}^{\text{sparse}} \odot \mathcal{X}) \tag{15}$$

$$\mathbf{z}^{\text{final}} = \sum_{i=1}^{N} \mathrm{w}_i \cdot \mathbf{z}_i = \text{sum}(\mathbf{w}^{\text{sparse}} \odot \mathcal{Z}), \tag{16}$$

where $i \in \{1, \ldots, N\}$, $\mathcal{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{B \times N \times K}$, and $\mathcal{X} = [\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_N] \in \mathbb{R}^{B \times N \times CH \times H \times W}$.

The ANIL meta-learning procedure for SNR-agnostic signal technology recognition is described as follows: Let a task $\mathcal{T} \sim p(\mathcal{T})$ be sampled from a distribution over tasks, where each task $\mathcal{T}$ corresponds to a specific SNR level and its own support and query sets, denoted by $\mathcal{D}_{\mathcal{T}}^{\text{support}}$ and $\mathcal{D}_{\mathcal{T}}^{\text{query}}$ respectively. The meta-learner model is denoted as:

$$f_\theta(\mathbf{S}) = h_\theta(\phi(\mathbf{S})), \tag{17}$$

where ϕ is a shared, frozen feature extractor during inner-loop adaptation and h_θ is the classification head updated during task-specific adaptation. The inner-loop adaptation step for h_θ using the support set is defined as:

$$\theta'_{\mathcal{T}} = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}}^{\text{support}}(h_\theta(\phi(\mathbf{S})), \mathrm{y}), \quad (\mathbf{S}, \mathrm{y}) \in \mathcal{D}_{\mathcal{T}}^{\text{support}}, \tag{18}$$

where α is the inner-loop learning rate and y are the corresponding labels. In the outer-loop meta-update adaptation across SNR levels, the model minimizes the loss on the query set to update θ as:

$$\min_\theta \mathbb{E}_{\mathcal{T} \sim p(\mathcal{T})} \left[\sum_{(\mathbf{S}, \mathrm{y}) \in \mathcal{D}_{\mathcal{T}}^{\text{query}}} \mathcal{L}_{\mathcal{T}}^{\text{query}}(h_{\theta'_{\mathcal{T}}}(\phi(\mathbf{S})), \mathrm{y}) \right]. \tag{19}$$

IV. ANALYSIS METHODOLOGY

A. Dataset and Experiment Setting

In this work, we use the publicly available LPWAN IQ signal dataset introduced by [24], captured using a universal software radio peripheral with a gain setting of 10 dB. Details of the signal acquisition process are provided in [24], and the noise injection methodology using MATLAB is described in [23]. Two groups of datasets are used in our experiments and are grouped by SNR. The first comprises 40,000 samples from eight known LPWAN signal classes (IEEE 802.15.4g, Sigfox, LoRa 7, 8, 9, 10, 11, and 12) across seven SNR levels (0 dB, -3dB, -6 dB, -10 dB, -13 dB, -16 dB, -20 dB). This dataset is used during meta-training and corresponds to signal conditions previously encountered during supervised training of the expert. The second group consists of 48,000 samples drawn from six previously unseen SNR levels (3 dB, 6 dB, 8 dB, -5 dB, -8 dB, -15 dB), with 8,000 samples per SNR and 1,000 samples per class. In other words, few-shot classification of LPWAN technologies across SNR levels on the first dataset with generalization to unseen SNRs on the second dataset.

B. Training Parameters

The meta-learner was trained using optimized hyperparameters, determined through the ray tune library [26]. The Adam optimizer is used for both inner and outer loop. The meta-learner uses a learning rate of $1e^{-3}$ with weight decay of $1e^{-4}$. We use 30% of the total samples of previously unseen dataset for the support set size. We perform 8-way (i.e. 8 classes) 5-shot adaptation over five independent runs, which means the number of labeled support examples per class that are available to learn from before classifying unseen queries.

C. Performance Metrics

To evaluate the performance of our model, we consider the confusion matrix, majorly the classification accuracy as the primary metrics. Furthermore, we assess computational efficiency at inference by analyzing the number parameters, memory usage, inference time and Floating-Point OPerations (FLOPs). The standard classification accuracy is computed as expressed in (20):

$$\text{Accuracy} = \frac{1}{N_s} \sum_{i=1}^{N_s} 1\left[\hat{\mathrm{y}}_i = \mathrm{y}_i\right], \tag{20}$$

where y_i is the ground truth label, $\hat{\mathrm{y}}_i$ is the predicted probability for class i, and N_s is the total number of samples. We compute FLOPs using the torchinfo library on a single 128×128 spectrogram input.

V. RESULTS

This section presents the results of the proposed model. We evaluate the performance under an 8-way 5-shot learning setting with sparse top-k values from 2 to 7. The results are obtained from our uncertainty-aware expert fusion with a meta-adaptive framework, averaged over five independent runs. Zero-Shot Learning or accuracy (ZSL) means how well can the model perform on unseen SNR levels without any adaptation while post-adaptation accuracy simply indicates how well it performs after few-shot adaptation using support data. Two approaches are compared: Meta-Adapter-based MoE and Sparse-Gated MoE. The Meta-Adapter framework introduces lightweight adapters inserted between the frozen denoising function and the classifier head, enabling adaptation by updating only the adapter parameters during meta-learning while keeping the experts unchanged. In contrast, the Sparse-Gated MoE use a gating mechanism to select and combine expert outputs but does not involve any further adaptation of the expert parameters.

Fig. 4a shows the overall average accuracy performance of different models for varying numbers of experts across unseen SNR levels. ANIL outperform other models. At low SNRs (e.g. -15 dB), Meta-Adapter achieves significant improvements, but ANIL achieves greater accuracy and remains more robust under severe noise conditions. As the SNR improves, ANIL continues to achieve the highest accuracy for most configurations when $k \geq 5$. Sparse-gated MoE shows moderate

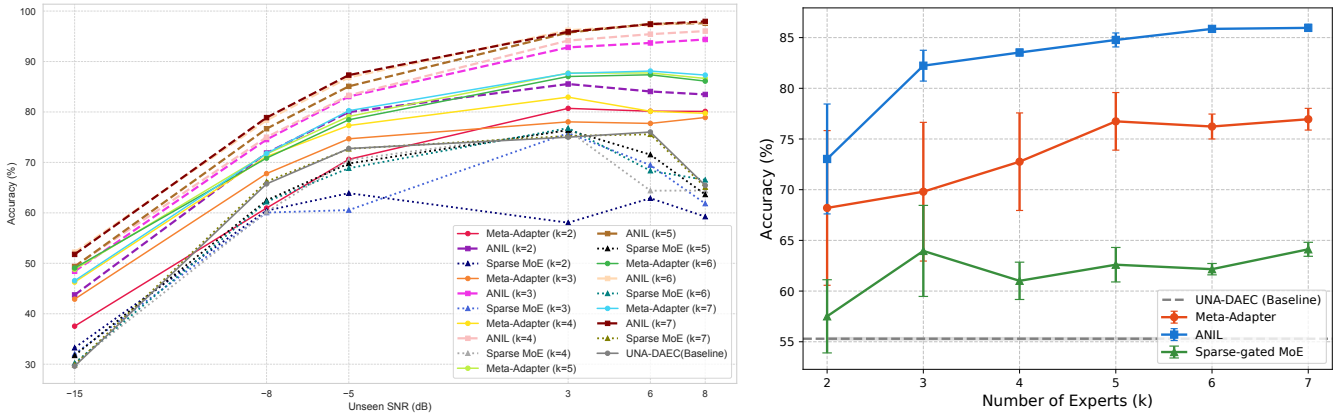

Fig. 4. (a) Overall average classification accuracy on unseen SNRs, and (b) Overall accuracy as a function of the number of experts after adaptation.

improvement over UNA-DAEC across certain SNR ranges. However, its performance fluctuates across different noise levels, making it less consistent than Meta-Adapter and ANIL. These results emphasize the effectiveness of post-adaptation strategies, particularly ANIL, in improving the generalisation of models to unseen SNR conditions.

Fig. 4b, shows the average performance accuracy and error bars indicate the corresponding standard deviation, based on number of experts. The baseline UNA-DAEC performance is also included as a horizontal dashed line for reference, along with its standard deviation band. Meta-Adapter and ANIL-based adaptation consistently outperform Sparse-Gated MoE across all values of k. The Meta-Adapter method shows a gradual improvement in accuracy as k increases, reaching its peak at k=7 with 76.95%, whereas ANIL achieves higher overall accuracy, peaking at k=7 with 85.96%. Notably, ANIL exhibits a steeper performance gain between k=2 and k=3, suggesting that additional experts are particularly beneficial in the early stages of adaptation. In contrast, Sparse-Gated MoE shows moderate improvements but with lower overall accuracy compared. The UNA-DAEC baseline achieves 55.30%, which is outperformed by both Meta-Adapter and ANIL methods even at lower k values, highlighting the effectiveness of post-adaptation in leveraging expert diversity. The relatively low standard deviation in ANIL for $k \geq 3$ indicates robust performance across runs. These findings highlight that adapting the classifier head significantly enhances classification performance with increasing expert models.

Fig. 5a presents the average performance gains of Meta-Adapter and ANIL methods when comparing post-adaptation results to their respective average zero-shot accuracy across varying numbers of experts (k). ANIL consistently demonstrates higher gains than Meta-Adapter for all values of k, reaching a maximum gain of approximately 20% at k=4. This indicates that fine-tuning the classifier head as done in ANIL is more effective for improving adaptation performance compared to adapter-based updates.

Fig. 5b compares the gains of the methods over the UNA-DAEC baseline. All methods surpass the baseline across all k values, with ANIL showing the largest gain of about 30% gain at k=6, and with optimal expert combination of k=4. The Sparse-Gated MoE achieves moderate gains but remains below others. Highlighting the best-performing number of experts for each method emphasizes that ANIL not only provides the highest accuracy gains but also does so consistently over multiple expert configurations. The inference time across all SNRs and k is 0.02 seconds for all models. This performance remains within a feasible range for real-time applications, balancing accuracy, computational efficiency and adaptiveness efficiently. Overall, these results emphasize the superior effectiveness and stability of the proposed model in leveraging multiple experts to achieve a substantial improvement in classification accuracy, generalization and adaptability for signal technology recognition tasks in dynamic wireless environments.

Fig. 5c compares our framework with SOTA models through our implementation, including the transformer-based Low SNR Recognition Model (LSRM) [17], MCformer [20], and Robust CNN (R-CNN) [22]. In particular, our ANIL-based variant outperforms all other models shown. This demonstrates the effectiveness of ANIL-based meta-learning in signal recognition.

A. *Computational Complexity*

Table I presents the model complexity in terms of memory usage, FLOPs, and total parameters. The ANIL-based MoE achieves the lowest memory footprint at 33.84 MB and moderate compute cost with 486.2 million FLOPs. It maintains 5.81 million parameters, all of which are frozen except for the classifier head, making it highly efficient for adaptation. In comparison, the Sparse Gated-MoE exhibits the highest complexity, requiring 41.31 MB of memory and 669.81 million FLOPs, since all weights and biases are updated during training. The Meta-Adapter-based MoE reduces FLOPs to 383.5 million but requires slightly more memory (34.76 MB) due to adapter layers. Other SOTA methods show varying trade-offs, but none match the efficiency of ANIL in achieving strong performance with minimal fine-tuning and competitive efficiency.

VI. CONCLUSION

We proposed a novel framework combining an uncertainty-aware gating mechanism, a sparse MoE architecture and ANIL meta-learning for enhanced signal technology recognition in

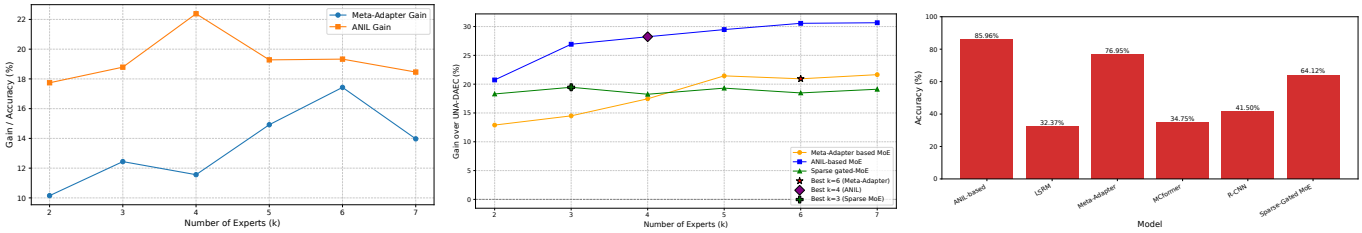

Fig. 5. (a) Accuracy gain over ZSL vs numbers of experts, (b) Gain over baseline and optimal number of experts, and (c) Comparison with SOTA models.

TABLE I
MODEL COMPLEXITY.

Model	Memory (MB)	FLOPs (M)	Total Parameters (M)
ANIL-based MoE	33.84	486.20	5.81
Meta-Adapter-based MoE	34.76	383.50	5.81
Sparse Gated-MoE	41.31	669.80	5.81
LSRM	26.61	186.53	4.87
MCformer	13.79	166.50	0.58
R-CNN	32.54	328.66	0.11

dynamic environments. This design enables more robust and efficient identification of signals and is suitable for deployment by regulators or operators on centralized servers, regional edge monitors or IoT gateways for spectrum monitoring. Experimental results demonstrated its strong performance in generalizing across unseen SNR levels and rapidly adapting via few-shot learning. While effective, future work could focus on enhancing computational efficiency through architectural or algorithmic optimizations, and exploring broader applicability in areas such as cross domain signal recognition, 5G signal analysis, and next-generation network monitoring.

REFERENCES

[1] IoT Analytics. (2024, Sep.) State of IoT Summer 2024: Number of connected IoT devices growing 13% to 18.8 billion globally. [Online; accessed 31-March-2025]. [Online]. Available: https://iot-analytics.com/number-connected-iot-devices/

[2] M. Islam, H. M. M. Jamil, S. A. Pranto, R. K. Das, A. Amin, and A. Khan, "Future industrial applications: Exploring lpwan-driven iot protocols," *Sensors*, vol. 24, no. 8, p. 2509, 2024.

[3] K. Mikhaylov, M. Stusek, P. Masek, R. Fujdiak, R. Mozny, S. Andreev, and J. Hosek, "Communication performance of a real-life wide-area low-power network based on sigfox technology," in *2020 IEEE International Conference on Communications (ICC)*. IEEE, 2020, pp. 1–6.

[4] M. Pérez, F. E. Sierra-Sánchez, F. Chaparro, D. M. Chaves, C.-I. Paez-Rueda, G. P. Galindo, and A. Fajardo, "Coverage and energy-efficiency experimental test performance for a comparative evaluation of unlicensed lpwan:Lorawan and Sigfox," *IEEE Acc*, vol. 10, pp. 97 183–97 196, 2022.

[5] P. Sun and G. Hu, "Robust identification algorithm of network communication signals via machine learning model," *Scalable Computing: Practice and Experience*, vol. 26, no. 2, pp. 552–565, 2025.

[6] E. Caushi, J. F. Lin, and R. Zhou, "Scalable 1-DFR-CNN for signal identification and classification over a wideband channel," in *IEEE 67th Int'l Midwest Symp. on Circuits and Systemss*, 2024, pp. 1335–1339.

[7] J. Chan, A. Wang, A. Krishnamurthy, and S. Gollakota, "DeepSense: Enabling carrier sense in low-power wide area networks using deep learning," *arXiv preprint arXiv:1904.10607*, 2019.

[8] W. Zhang, D. Huang, M. Zhou, J. Lin, and X. Wang, "Open-set signal recognition based on transformer and Wasserstein distance," *Applied Sciences*, vol. 13, no. 4, p. 2151, 2023.

[9] Z. Wang, "Signal recognition model based on transformer," *Academic Journal of Eng. and Tech. Science*, vol. 7, no. 4, pp. 187–195, 2024.

[10] M. Akrout, A. Mezghani, E. Hossain, F. Bellili, and R. W. Heath, "From multilayer perceptron to gpt: A reflection on deep learning research for wireless physical layer," *IEEE Comm. Mag.*, vol. 62,no.7,pp.34-41, 2024.

[11] A. Vats, R. Raja, V. Jain, and A. Chadha, "The evolution of mixture of experts: A survey from basics to breakthroughs," *Preprint. (Aug.)*, 2024.

[12] Z. Feng, Z. Zhang, X. Yu, Y. Fang, L. Li, X. Chen, Y. Lu, J. Liu, W. Yin, S. Feng, Y. Sun, L. Chen, H. Tian, H. Wu, and H. Wang, "ERNIE-ViLG 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 10 135–10 145.

[13] V. Negroni, D. Salvi, A. I. Mezza, P. Bestagini, and S. Tubaro, "Leveraging mixture of experts for improved speech deepfake detection," *arXiv preprint arXiv:2409.16077*, 2024.

[14] Y. Balaji, S. Nah, X. Huang, A. Vahdat, J. Song, Q. Zhang, K. Kreis, M. Aittala, T. Aila, S. Laine, B. Catanzaro, T. Karras, and M.-Y. Liu, "eDiff-I: Text-to-image diffusion models with an ensemble of expert denoisers," 2023. [Online]. Available: https://arxiv.org/abs/2211.01324

[15] Y. Li, Y. Zhang, C. Wang, Z. Zhong, Y. Chen, R. Chu, S. Liu, and J. Jia, "Mini-Gemini: Mining the potential of multi-modality vision language models," *arXiv preprint arXiv:2403.18814*, 2024.

[16] J. Guo, Y. Cai, K. Bi, Y. Fan, W. Chen, R. Zhang, and X. Cheng, "CAME: Competitively learning a mixture-of-experts model for first-stage retrieval," *ACM Trans. on Inf Syst*, vol. 43, no. 2, pp. 1–25, 2025.

[17] J. Gao, Q. Cao, and Y. Chen, "MoE-AMC: Enhancing automatic modulation classification performance using mixture-of-experts," *arXiv preprint arXiv:2312.02298*, 2023.

[18] L. Ilias, G. Doukas, V. Lamprou, C. Ntanos, and D. Askounis, "Convolutional neural networks and mixture of experts for intrusion detection in 5G networks and beyond," *arXiv preprint arXiv:2412.03483*, 2024.

[19] N. Shazeer, A. Mirhoseini, K. Maziarz, A. Davis, Q. Le, G. Hinton, and J. Dean, "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer," *arXiv preprint arXiv:1701.06538*, 2017.

[20] S. Hamidi-Rad and S. Jain, "Mcformer: A transformer based deep neural network for automatic modulation classification," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[21] M. Z. Nisar, M. S. Ibrahim, M. Usman, and J.-A. Lee, "A lightweight deep learning model for automatic modulation classification using residual learning and squeeze–excitation blocks," *Applied Sciences*, vol. 13, no. 8, p. 5145, 2023.

[22] O. F. Abd-Elaziz, M. Abdalla, and R. A. Elsayed, "Deep learning-based automatic modulation classification using robust cnn architecture for cognitive radio networks," *Sensors*, vol. 23, no. 23, p. 9467, 2023.

[23] I. O. Ebo, I. Ajayi, L. Mroueh, Y. Ziade, and S. Azarian, "A channel attention-based denoising autoencoder u-net and classifier (una-daec) for signal identification under challenging snr conditions," *IEEE Access*, vol. 13, pp. 77 317–77 333, 2025.

[24] A. Shahid, J. Fontaine, M. Camelo, J. Haxhibeqiri, M. Saelens, Z. Khan, I. Moerman, and E. De Poorter, "A convolutional neural network approach for classification of LPWAN technologies: Sigfox, LoRA and IEEE 802.15.4g," in *16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2019, pp. 1–8.

[25] A. Raghu, M. Raghu, S. Bengio, and O. Vinyals, "Rapid learning or feature reuse? towards understanding the effectiveness of maml," *arXiv preprint arXiv:1909.09157*, 2019.

[26] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," *arXiv preprint arXiv:1807.05118*, 2018.