Meta-Learning for Wireless Interference Identification

Ali Owfi*, Fatemeh Afghah*, Jonathan Ashdown †
*Holcombe Department of Electrical and Computer Engineering, Clemson University, Clemson, SC, USA
Emails: aowfi@clemson.edu, fafghah@clemson.edu
†Air Force Research Laboratory, Rome, NY 13441, USA
Email: jonathan.ashdown@us.af.mil

2023 IEEE Wireless Communications and Networking Conference (WCNC) | 978-1-6654-9122-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCNC55385.2023.10119039

I. Abstract

Deep learning-based (DL-based) models have shown to be powerful tools for wireless interference identification (WII). However, one of the key concerns toward using these models in practical systems is that they perform poorly when they are encountered with signals coming from new sources not previously observed during the training phase. In a real-world communication system, the interference identifier will frequently face new unknown signals due to the existence of many wireless transmitters. This renders the conventional DL-based models impractical as a WII tool unless they go through a new training phase. Retraining the model is not only inefficient, but it can also be not feasible in some cases (e.g., at end-user devices) as the training phase consumes time and resources and requires large amounts of data. We present a new approach for data-driven WII systems using meta- learning to address the lack of adaptability in conventional DL-based models to new (not previously seen) signals. We show that by using meta-learning, we are able to identify signals coming from not previously observed technologies and frequencies using just a handful of new samples, a task that is not generally possible with conventional DL models. Finally, we analyze and compare the performance of the presented meta-learning model in multiple different settings using raw I/Q samples and Fast Fourier Transform of I/Q samples. Based on our experiments, we show that the proposed meta-learning scheme outperforms the conventional deep learning models for WII when there are just a few samples available for training[1].

II. Introduction

Traditionally, the physical layer has been grounded upon solid mathematics and information theory. In many cases, the methods proposed based upon these pillars are theoretically proven to be optimal based on a few prior assumptions that define the effective parameters. The issue is that, in practical settings, model-based approaches may fail in accurately modeling some system components and parameters such as channels and hardware impairments. Thus, these methods deem ineffective in new environments in the presence of unknown channel models or other sources of uncertainties such as time and frequency offsets. Moreover, these optimization-driven model-based solutions can suffer from high computational complexity.

In these non-linear and dynamic settings where we can not correctly capture all details theoretically with mathematical models, data-driven approaches and, more specifically, neural network models can be potential solutions as they have been shown to be universal function approximators [1]. While neural network models are not mathematically traceable in general, they can capture the non-linearities that lie in physical layer context to a reasonable extent. Neural network models have been recently utilized in a variety of well-known problems in physical layer-related applications such as autoencoder-based waveform design, and modulation detection [2], [3], [4], [5].

Interference in communication systems can be caused by various sources, including non-cooperative coexisting technologies, communication devices operating in adjacent channels, or jammers. Identifying sources of interference signals which is the goal of wireless interference identification (WII) is a key step of interference management strategies for coexistence technologies. WII also has an essential role in detecting intentional radio frequency interference (RFI) or jamming to enhance the security of communication networks [6]. Classical machine learning techniques which are built upon knowledge-based extracted features such as transform-based features (e.g., short-time Fourier transform) or statistical features have been utilized for interference detection, and classification in various communication systems [7]. However, these methods fail to offer the expected interference detection and identification accuracy as they rely on expert knowledge and engineered features.

Deep learning approaches with the capability of automated feature learning and latent representation from raw data have shown significant performance in wireless interference detection in different communication systems. In recent works, the performance of multiple neural networks architectures such as convolutional neural networks (CNN) and long short-term memory (LSTM) for WII have been investigated in unlicensed frequency bands such as the 2.4 GHz [8], [9]. However, these deep learning-based models suffer from a lack of generalization. That means when a conventional DL-based model is trained on a specific dataset, it often achieves a relatively high accuracy on the corresponding test set. However, if the classifier is encountered with new data collected from different transmitters or a new environment, it will fail to perform accurate classification unless it goes through a training phase again. The re-training phase indeed is a costly process in terms of required time and computational resources. Taking into account that in real-world situations, the classifier will encounter waveforms from unknown transmitters not observed

[1]This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-20-1-0090 and the National Science Foundation under Grant Numbers ECCS-2202972. Approved for Public Release; Distribution Unlimited: AFRL-2022-3094.

before in the training dataset, we can see that this lack of adaptability in such classifiers is a critical issue, and if left unaddressed, it can limit the application of DL-based WII models in real-world applications.

In this paper, we address the problem of lack of adaptability, which exists in a majority of existing DL-based WII models, by developing a meta learning-based WII technique for coexistence management in the 2.4 GHz ISM band using the dataset collected in [9]. In meta-learning, instead of learning a single task with a specific dataset, the aim is to learn the learning process itself. In other words, the model learns how to adapt quickly to a new task with a new dataset. To talk in the context of transmitter classification in wireless interference identification, a model trained using meta-learning would be able to classify among transmitters that have not been observed before during the offline training. This will make this model more adaptable in real-world scenarios compared to conventional DL-based models. Therefore, the main contribution of this paper is to develop a meta-learning model for cross-technology wireless interference identification to train a model which is adaptable to tasks and classes from technologies that have not been observed previously, just by observing a few samples for each new task. To the best of our knowledge, no prior work has used meta-learning for classification tasks in cross-technology wireless interference identification. This use of meta-learning will let the classifiers reach a better generalization point and potentially make Dl-based classification approaches in the physical layer viable in real-world scenarios where the models need to be highly adaptable.

The rest of this paper is structured as follows: Section 3 discusses the previous works on DL-based classification approaches in the context of physical layer problems and wireless interference identification. In Section 4, a brief introduction to meta-learning and model agnostic meta-learning approach is provided, and then our proposed meta-learning-based methodology to solve the issue of adaptability in DL-based WII is described. In Section 5, the scenarios used to evaluate the performance of the proposed meta-learning model and their results are discussed, followed by concluding remarks in Section 6.

III. Related Work

In [8], a CNN-based model was used for classifying ten different single carrier modulation schemes based on their radio frequency time-series data. The CNN-based model achieved competitive results with respect to a few other traditional schemes used for modulation classification, slightly beating them in terms of classification accuracy both in high and low SNRs.

The authors in [9] proposed a CNN-based model for wireless interference identification, where the model was trained to classify among 15 classes of each denoting a different channel from 3 technologies of IEEE 802.11 b/g, IEEE 802.15.4, and IEEE 802.15.1, using raw I/Q samples. The identification task is performed based on 128 I/Q Samples with a duration of 12.8 μs on a 10MHz bandwidth. Building upon [9], in [10] other DL-based models such as LSTM, ResNet, and Convolutional Long- Short Term Deep Neural Network (CLDNN) were tested on the same dataset, all giving relatively high classification accuracy. They also focused on decreasing the training-phase time for the CNN model, achieving a model which is trained in approximately 60% of the previous CNN-based model training time while giving comparable accuracy.

In [11], the authors focused on improving the adaptability of current DL-based models for WII by decreasing the calculation cost of the neural network model and thus, decreasing the training time. They divided a CNN model into several blocks, each having its output. The model is designed such that based on the difficulty of the given output, an appropriate depth of blocks is chosen so that unnecessary resources are not allocated.

While the existing works around solving classification tasks in the context of physical layer using neural networks have been promising, current models lack adaptability in the sense that they can perform well on the tasks that they were trained on and the classes that they have observed during training. However, they will perform poorly if they are encountered with unknown classes and new classification tasks unless they go through training again. This issue makes these models unsuitable in real-world scenarios where the classifier may receive signals from sources with unknown or new modulation schemes or technologies. This concern for adaptability has been previously addressed using meta-learning in physical layer. In [12], an introduction to meta-learning methods with the application to communication systems is given. Also, two specific use cases of meta-learning in supervised learning for demodulation and unsupervised learning for transmission and reception are discussed. In [13], instead of training an autoencoder on a fixed channel or a fixed set of channels similar to the prior works, an autoencoder is trained using meta-learning with the goal of adaptability so that it becomes applicable to different types of channels rather than being applicable just to a fixed type of channel. Furthermore, in [14], a meta-learning approach is proposed for specific emitter identification. However, to the best of our knowledge, no existing work has used meta-learning to address the issue of unadaptability for supervised classification tasks that are defined in the context of cross-technology interference identification.

IV. Methodology

Our goal in this paper is to address the lack of adaptability in the existing DL-based WII models. We intend to develop a DL-based WII that can adapt quickly to new tasks and classes that have never been seen before during the training phase using a meta-learning approach. In this section, the concept of meta-learning, the meta-learning model that we used, and how we defined a meta-learning problem in the context of WII are discussed.

A. Meta Learning

Conventional deep learning models are designed to handle one task with an objective (i.e., a supervised learning problem) over a dataset which is divided into a training set and a testing set. After training the model on the training set, the model is expected to perform relatively well on the test set too. In contrast, in meta-learning or also popularly known as 'learning to learn', the objective is to improve a learning model (known as the base learner/learner/inner model) such that it can adapt to new tasks quickly while using a limited amount of data. In other words, the objective is to learn the learning process itself. Contrary to conventional deep learning models, which in each iteration, we observe a batch of training samples, in meta-learning, we observe a batch of new tasks, each with a different objective, a train set (known as *support set*), and a test set (known as *query set*). For each task, the model trains on the corresponding support set, learning to do the task at hand. This is commonly known as the adaptation phase or the inner loop. After that, the loss function on that task is calculated using the query set. The necessary changes to the model are then applied based on the calculated loss of the tasks and the outer objective of the meta-learning model. This part is known as the outer loop. It should be noted that the outer objective function in a meta-learning model is different from the inner objective function of each task.

B. Model-Agnostic Meta-Learning (MAML)

Multiple algorithms have been proposed to accomplish the goal of meta-learning, which is to provide a model that is able to adapt to new tasks fairly quickly. Model-Agnostic Meta-Learning (MAML) [15] is one of such algorithms. A general block diagram of MAML is shown in Figure 1. In MAML, the idea is to find an initial set of parameters for the inner model from which adapting to new tasks would be as fast as possible. To define this process formally, MAML considers an inner model f with a set of parameters θ denoted by f_θ.

In the inner loop, for each adaptation phase to a new task $\mathcal{T}_i$, the model parameters θ are updated to θ' (steps 1, 2 and 3 in Figure 1). The equation below shows this updating phase if one gradient step is taken, but it can be generalized to cases where multiple gradient steps are taken as well.

$$\theta_i^{'} = \theta - \alpha \Delta_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta) \tag{1}$$

where α is the step size.

In the outer loop, the objective function is defined as below:

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta_i'}\right) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta - \alpha\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)}\right) \tag{2}$$

Where we aim to optimize f_θ' with respect to θ which is the initial set of parameters that the inner model uses to adapt to each task.

And finally the outer loop optimization (meta-optimization) is formulated as follows (steps 4, 5 and 6 in Figure 1):

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}\left(f_{\theta_i'}\right) \tag{3}$$

where β is a hyper-parameter known as *meta-step size*.

During the meta-testing phase, for each task, the adaptation is performed, and the loss function on that task is calculated, but θ or the inner model's initial set of parameters is not updated in contrast to the meta-training phase.

C. Proposed Meta learning-based WII Model

The first step toward forming a meta-learning model is to define the tasks that the model needs to tackle. We define each task to be a classification task among n different classes where for each class, there are a limited number of samples that the model can train on, denoted by k. More formally, this problem is called an N-way k-shot classification.

Moreover, we divide the classes into two groups, where one group is used to create tasks for the meta-training phase, and the other group is used to create tasks for the meta-testing phase. This way, we will be able to test the extent of our model's adaptability as the classes in meta-testing tasks will be unknown to the model. The proposed WII model is developed based on the MAML method described in Section IV-B, where a CNN model with 3 convolution layers and 2 dense layers is used in the inner model. Details of the inner model structure can be seen in Table I and details about parameters are shown in II.

TABLE I
STRUCTURE OF INNER MODEL

Layer	Input	Parameters	Activision Function
1D Convolution	2 * 128	kernel size = 3 64 out channels stride = 1 padding =1	ReLU
1D Average Pooling	64*128	kernel size = 3	-
1D Convolution	64 * 42	kernel size = 3 512 out channels stride = 1 padding =1	ReLU
1D Average Pooling	512* 42	kernel size = 3	-
1D Convolution	512 * 14	kernel size = 3 1024 out channels stride = 1 padding =1	ReLU
1D Average Pooling	1024 * 14	kernel size = 3	-
Dense Layer	4096	128 neurons	ReLU
Dense Layer	128	4 neurons	Softmax

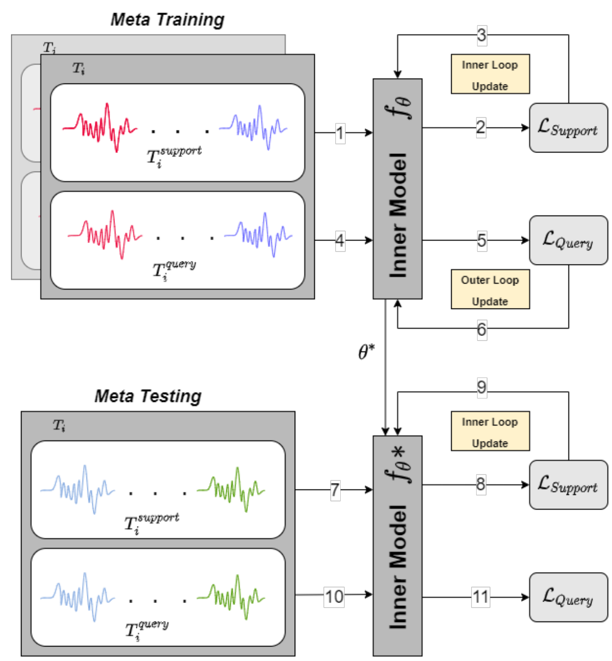


Fig. 1. *Block diagram of MAML. Numbers 1-6 and 7-11 correspond to the steps taken in meta-training and meta-testing, respectively.*

V. EXPERIMENTS AND RESULTS

A. Dataset

In this study, we used the dataset generated by an SMBV100A vector signal generator in [9] that consists of sample vectors of 128 complex-valued baseband I/Q samples corresponding to 15 different classes. These classes represent packet transmissions of IEEE 802.11 b/g, IEEE 802.15.4, and

TABLE II
MODEL PARAMETERS

Parameter	Value
Loss Function	CrossEntropy
Optimizer	Adam
Inner Step Size	0.02
Outer Step Size	0.001
Outer Decay Rate	0.97 per 10 Epochs
# Inner Adaptations	5
Batch Size	48

IEEE 802.15.1 spanning different physical layer specifications with overlapping frequency channels within the 2.4 GHz ISM band. Information about the classes and their corresponding technology and frequency can be seen in Table III. The sample vectors for each class are captured from an SNR range of -20 dB to 20 dB with a step size of 2 dB and there are a total of 715 sample vectors for each class in each SNR. These I/Q samples have also been transformed to the frequency domain using Fast Fourier Transform (FFT) and so the frequency domain data of these samples are also available [16].

TABLE III
DESCRIPTION OF 15 CLASSES USED IN THE DATASET.

Class ID	Technology	Center Frequency	Channel Width
0	IEEE 802.15.1	2422 MHz	1 MHz
1	IEEE 802.15.1	2423 MHz	1 MHz
2	IEEE 802.15.1	2424 MHz	1 MHz
3	IEEE 802.15.1	2425 MHz	1 MHz
4	IEEE 802.15.1	2426 MHz	1 MHz
5	IEEE 802.15.1	2427 MHz	1 MHz
6	IEEE 802.15.1	2428 MHz	1 MHz
7	IEEE 802.15.1	2429 MHz	1 MHz
8	IEEE 802.15.1	2430 MHz	1 MHz
9	IEEE 802.15.1	2431 MHz	1 MHz
10	IEEE 802.11 b/g	2422 MHz	20 MHz
11	IEEE 802.11 b/g	2427 MHz	20 MHz
12	IEEE 802.11 b/g	2432 MHz	20 MHz
13	IEEE 802.15.4	2425 MHz	2 MHz
14	IEEE 802.15.4	2430 MHz	2 MHz

B. Experimental Scenarios

For our experiments, we utilize both the raw I/Q samples and their FFT transformations available in [9]. Moreover, for each type of input, we define two scenarios which differ in the classes that are excluded from the meta-training tasks. The two scenarios are as follows:

- Scenario 1: The experiment is designed as a 4-way n-shot classification, where 4 denotes the number of classes in each task and n represents number of samples for each class. Five out of 15 classes are excluded from the meta-training tasks and for meta-testing tasks, only those five unseen classes are used. These five classes are selected from all of the three technologies that are available in the dataset, meaning that while the model is encountered with classification tasks on unseen classes, those unseen classes are not corresponding to entirely unknown technologies.
- Scenario 2: The experiment is designed as a 4-way n-shot classification. Five classes are excluded from the meta-training tasks and for meta-testing tasks, only those 5 unseen classes are used. These five selected classes are the only classes which represent IEEE 802.15.4 and IEEE 802.11 b/g that are excluded from the training tasks, meaning that the model will be tested by classification tasks among classes representing completely unknown technologies to the model.

TABLE IV
ACCURACY OF META-LEARNING MODEL WITH 95% CONFIDENCE INTERVAL IN MULTIPLE SETTINGS.

1 SHOT

	IQ		FFT	
Scenario	1	2	1	2
High SNR	97.22 ± 0.25	78.53 ± 0.56	95.62 ± 0.47	78.15 ± 0.55
Mid SNR	89.23 ± 0.43	73.18 ± 0.61	94.76 ± 0.36	80.30 ± 0.49
Low SNR	30.11 ± 0.67	39.15 ± 0.64	38.19 ± 0.68	40.19 ± 0.66

2 SHOT

	IQ		FFT	
Scenario	1	2	1	2
High SNR	98.45± 0.13	81.10 ± 0.45	98.85 ± 0.12	85.02 ± 0.35
Mid SNR	95.63 ± 0.26	77.56 ± 0.74	96.67 ± 0.25	82.79 ± 0.38
Low SNR	32.58 ± 0.46	42.54 ± 0.48	44.47 ± 0.50	43.15 ± 0.52

5 SHOT

	IQ		FFT	
Scenario	1	2	1	2
High SNR	99.35 ± 0.04	87.28 ± 0.30	99.35 ± 0.05	88.98 ± 0.32
Mid SNR	96.79 ± 0.12	83.54 ± 0.32	98.82 ± 0.12	86.11 ± 0.28
Low SNR	36.20 ± 0.31	44.44 ± 0.36	54.64 ± 0.35	46.27 ±0.37

Multiple reasons led us to define the tasks in both scenarios as a 4-way classification. Firstly, while conventional models often benefit from having large amounts of training data, meta-learning models benefit from large amounts of different tasks to train on so that they can become well-generalized to distinctive tasks. To show the capability of the meta-learning model in identification of new classes with a few samples, we set aside one third of available classes (i.e., 5 classes) for the meta-testing tasks, thus left with 10 classes for the meta-training tasks. In this case, choosing 5-way classification tasks would have maximized the number of different tasks we would observe in meta-training ($\binom{10}{5}$=252), however, it would have resulted in just one distinctive task for meta-testing ($\binom{5}{5}$=1), which is not desirable. By choosing a 4-way classification, we ended up with having $\binom{10}{4} = 210$ different meta-training tasks and $\binom{5}{4} = 5$ different meta-testing tasks.

C. Results

Table IV indicates the classification accuracies of our meta-learning model during the meta-testing phase in different settings. The model was trained on data corresponding to 3 SNR ranges using either the FFT data or the I/Q sample data from [16]. High SNR, mid SNR, and low SNR correspond to data in SNR ranges of -20 to -14 dB, -2 to 4 dB, and 14 to 20 dB, respectively. Moreover, for each of these six combinations, the two scenarios explained earlier were tested using either 1, 2 or 5 samples per class to train the meta-learning model. It should be emphasized that the classes in meta-testing tasks did not overlap with the classes used in meta-training tasks.

A few observations can be made from this Table. Firstly, the meta-learning model performed better in the first scenario compared to the second scenario. This is logical as in scenario 2, the model had to classify between classes corresponding to not previously observed communication technologies (IEEE 802.15.4 and IEEE 802.11 b/g technologies were not seen in the training set), but in scenario 1, none of the communication technologies were new to the model. Thus, it seems that the model was able to reach a better generalization for all the classes of all three technologies in scenario 1. It can also be concluded from Table IV that the model will perform better when the model is trained on input data from high SNR ranges. This result was also observed for conventional deep learning models used in WII and modulation classification works [4], [9], [11]. Furthermore, it can be observed that the higher the number of shots, the higher the accuracy will be, which is logical.

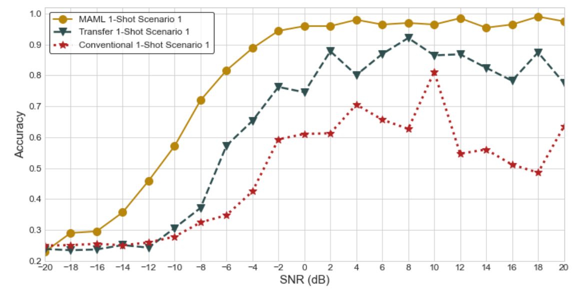
Regarding the performance of the model when different input data types are used (i.e., raw IQ samples or the FFT transformations), no strong statements can be given to pinpoint one as superior over the other. Nevertheless, it is worth noting that the model showed better robustness to high noise levels when FFT data were used. This can be seen in table IV, where FFT cases have higher accuracy in low SNR ranges than their IQ counterparts.

To denote the higher adaptability of meta-learning over conventional deep learning in the context of WII, a comparison has to be made between the two. A meta-learning model goes through several sets of training on different tasks during the meta-training phases. Then it has to adapt itself to a new task created from unseen classes during the meta-testing phase by training on only a handful of samples from the corresponding task. To have a fair comparison, we consider a conventional model already trained on a specific task, that has to adapt to a new task by training on only a handful of samples from the new task. This can be regarded as transfer learning. Furthermore, we will also consider a conventional model without any prior offline training that has to adapt to a new task by also training on just a few samples.

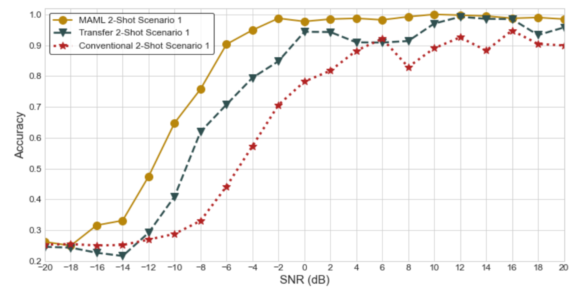
Figures 2 and 3 compare the accuracy of the meta-learning model (MAML), the conventional deep learning model with prior training (denoted as transfer learning), and the conventional deep learning model without prior training in different SNRs. The model structure used for the conventional deep learning and transfer learning is the same as the inner model used for the MAML model (model structure can be seen in table I). The results seen in Figures 2 and 3 correspond to results based on scenario one and scenario two, respectively, which were explained previously. Each sub-plot in figures 2 and 3 indicates the results when a different number of samples were given to the models to train on (one, two, and five samples per class, i.e., 1-shot, 2-shot, and 5-shot classification). It should be noted once again, that none of the tasks and classes observed in these experiments were previously observed by the MAML model during the meta-training phase.

As it can be seen from figures 2 and 3, it is clear that for every case, MAML outperforms the conventional models with prior training (transfer learning) and without prior training in terms of accuracy. As expected, all models have higher accuracies in 5-shot cases compared to 2-shot and 1-shot cases. However, the conventional models are drastically hindered in cases where less data is available. The difference in the conventional models' performances between the 1-shot cases and the 5-shot cases are larger compared to the same difference in performance for MAML. In other words, MAML can learn to do a new task way more efficiently than the conventional model by using less data while giving higher accuracy since it has learned to learn new tasks quickly during its meta-training phase. This denotes the adaptability of meta-learning, which is desired for WII in a real-world situation where the model has to adapt to new tasks quickly.

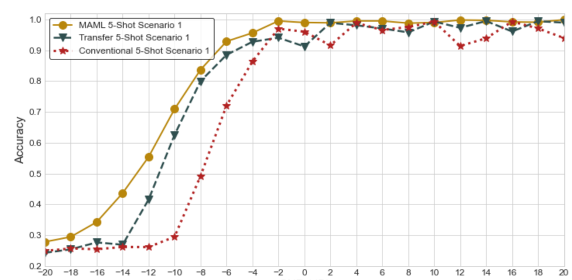
Furthermore, since just a handful of samples were used to train the models, the models' performances depend on how well those few samples represent the test set. Thus, randomness is higher when less data is used to train the models, and so, we can observe fluctuations for all models to some extent in figures 2 and 3. Nevertheless, MAML seems to be much more stable in all cases as these fluctuations are way more visible for the conventional model and the transfer learning model, especially in 1-shot results. This stability could be an indicator of the generalization that MAML has reached during its meta-training phase.



(a) 1 Shot

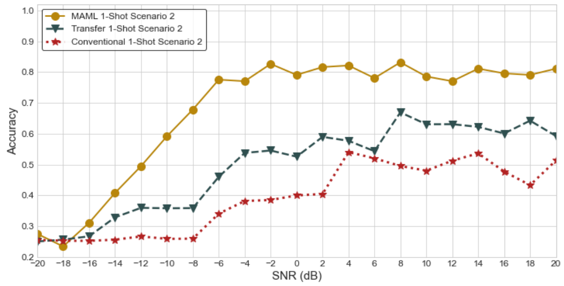


(b) 2 Shot

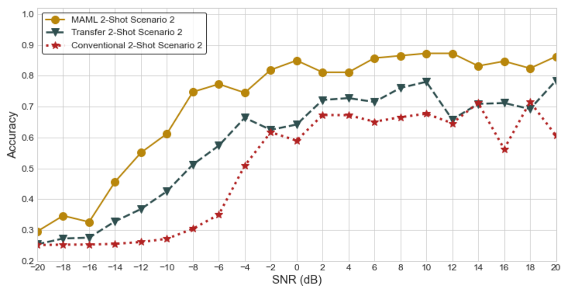


(c) 5 Shot

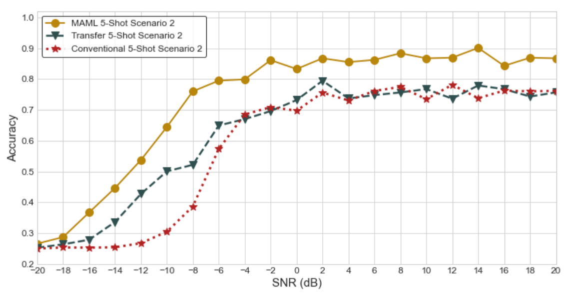
Fig. 2. Accuracy of the trained meta-learning model (MAML) compared to conventional DL model with and without prior learning based on scenario 1. Sub-figures a,b and c represent cases, where the number of samples per class used to train the models (number of shots) were one, two, and five, respectively.



(a) 1 Shot



(b) 2 Shot



(c) 5 Shot

Fig. 3. Accuracy of the trained meta-learning model (MAML) compared to conventional DL model with and without prior learning based on scenario 2. Sub-figures a,b and c represent cases, where the number of samples per class used to train the models (number of shots) were one, two, and five, respectively.

VI. Conclusion

This paper addressed the lack of adaptability that exists in current DL-based approaches in wireless interference identification by utilizing meta-learning.

To mimic a practical situation, we considered a coexistence system where the interference identifier has to adapt to new signals from unknown technologies and frequencies using just a handful of samples. To evaluate the impact of a meta-learning-based WII solution, we compared the performance of a meta-learning model against a conventional DL model in the coexistent system explained above. The dataset we used included data from three different wireless technologies (IEEE 802.15.1, IEEE 802.11 b/g, and IEEE 802.15.4) operating in frequencies from 2422 MHz to 2430 MHz. Our findings show that we can significantly outperform conventional deep learning for WII by using meta-learning, especially when there are just a few samples for the models to train on. Moreover, even though training on just a few samples introduces high randomness to the DL models in general, we showed that meta-learning would be less impacted by that randomness as it reaches higher stability in results.

VII. Acknowledgement and Disclaimer

This work is funded in part by the Air Force Office of Scientific Research (AFOSR) contract number FA9550-21-1-0314 and the National Science Foundation under Grant Numbers CNS-2202972. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Government.

References

[1] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[2] T. J. O'Shea, K. Karra, and T. C. Clancy, "Learning to communicate: Channel auto-encoders, domain specific regularizers, and attention," in *2016 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, 2016, pp. 223–228.

[3] S. Xue, Y. Ma, N. Yi, and R. Tafazolli, "Unsupervised deep learning for mu-simo joint transmitter and noncoherent receiver design," *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 177–180, 2019.

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

[5] T. Erpek, T. J. O'Shea, and T. C. Clancy, "Learning a physical layer scheme for the mimo interference channel," *2018 IEEE International Conference on Communications (ICC)*, pp. 1–5, 2018.

[6] S. Ujan, N. Navidi, and R. J. Landry, "Hierarchical classification method for radio frequency interference recognition and characterization in satcom," *Applied Sciences*, vol. 10, no. 13, 2020.

[7] G.-S. Wang, Q.-H. Ren, and Y.-Z. Su, "The interference classification and recognition based on sf-svm algorithm," in *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, 2017, pp. 835–841.

[8] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[9] M. Schmidt, D. Block, and U. Meier, "Wireless interference identification with convolutional neural networks," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE, 2017, pp. 180–185.

[10] X. Zhang, T. Seyfi, S. Ju, S. Ramjee, A. El Gamal, and Y. C. Eldar, "Deep learning for interference identification: Band, training snr, and sample selection," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.

[11] P. Wang, Y. Cheng, and B. Dong, "Multi-depth adaptive networks for wireless interference identification," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[12] O. Simeone, S. Park, and J. Kang, "From learning to meta-learning: Reduced training overhead and complexity for communication systems," in *2020 2nd 6G Wireless Summit (6G SUMMIT)*. IEEE, 2020, pp. 1–5.

[13] S. Park, O. Simeone, and J. Kang, "Meta-learning to communicate: Fast end-to-end training for fading channels," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5075–5079.

[14] N. Yang, B. Zhang, G. Ding, Y. Wei, G. Wei, J. Wang, and D. Guo, "Specific emitter identification with limited samples: a model-agnostic meta-learning approach," *IEEE Communications Letters*, vol. 26, no. 2, pp. 345–349, 2021.

[15] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 1126–1135.

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