

Radio Frequency Interference Detection for SMAP Radiometer Using Convolutional Neural Networks

Ahmed Manavi Alam[✉], *Graduate Student Member, IEEE*, Mehmet Kurum[✉], *Senior Member, IEEE*, and Ali C. Gurbuz[✉], *Senior Member, IEEE*

Abstract—Passive remote sensing is a crucial technology for climate studies and Earth science. National Aeronautics and Space Administration’s soil moisture active passive (SMAP) is a remote sensing observatory that uses passive microwave radiometer measurements to estimate soil moisture and detect the freeze or thaw state. Despite operating in the protected band of the radio spectrum (1400–1427 MHz), the radiometer’s measurements are nonetheless tainted by radio frequency interference (RFI). An increasing number of radio frequency transmissions such as those from air surveillance radars, 5G wireless communications, and unmanned aerial vehicles are contributing to RFI through either out-of-band emissions or operating in-band illegally. Physical modeling to detect RFI globally might prove to be challenging as RFI can be generated from single as well as multiple sources and these can be divided as pulsed or continuous wave RFI. In this study, a deep learning (DL) based RFI detection method is proposed with a novel convolutional neural network framework that can detect different types of RFI on a global scale. This is a data-driven approach where the detection framework learns directly from the SMAP data products to make a decision whether a certain footprint is RFI contaminated or not. SMAP’s level 1 A data products containing antenna counts of different raw moments along with Stokes parameters are used in this study to produce spectrograms and level 1B data products containing the quality flags are used to dynamically label those spectrograms. This study’s robust DL framework provided the highest accuracy with the raw moments of horizontal polarization (99.99%) to detect RFI globally.

Index Terms—Deep learning (DL), radio frequency interference (RFI), remote sensing, soil moisture active passive (SMAP).

I. INTRODUCTION

RADIO frequency interference (RFI) has become an important issue for both active and passive remote sensing. To remotely assess the features of the Earth’s surface and atmosphere, passive microwave remote sensing makes use of natural thermal emissions [1]. Because of its sensitivity to a specific attribute of interest and less attenuation by the intervening atmosphere between the source of emission and the sensor, the microwave section of the electromagnetic spectrum is frequently well suited

Manuscript received 17 June 2022; revised 23 September 2022 and 14 November 2022; accepted 15 November 2022. Date of publication 18 November 2022; date of current version 30 November 2022. This work was supported by the National Science Foundation under Grant 2030291 and Grant 2047771. (Corresponding author: Ali C. Gurbuz.)

The authors are with the Department of Electrical and Computer Engineering, and Information Processing and Sensing Lab, Mississippi State University, Mississippi State, MS 39672 USA (e-mail: aa2863@mssstate.edu; kurum@ece.mssstate.edu; gurbuz@ece.mssstate.edu).

Digital Object Identifier 10.1109/JSTARS.2022.3223198

for this purpose. However, the microwave region’s relative insensitivity to atmospheric phenomena makes it a particularly appealing spectral range for wireless communication and radars. This increasing demand from communications has the potential to lead to RFI that increases the noise floor which deteriorates the performance and products of remote sensing systems. The allotted spectrum for microwave remote sensing instruments can fall victim to interference from neighboring wireless systems, including the deployment of 5G accompanied by unsanctioned communication devices.

The growing demand for bandwidth in commercial applications necessitates the development of coexistence techniques between passive radiometry and future wireless infrastructure. To study such problem, we initiated the development of a unique physical testbed for collecting remote sensing datasets with ground truth in the presence of communication signals [2]. As a part of such effort, we chose the National Aeronautics and Space Administration’s soil moisture active passive (SMAP) satellite mission to develop initial learning-based RFI detection models. Once the testbed is complete, it will enable training, optimization, and benchmarking of such models, possibly for mitigation. While this article focuses on the development of learning models for SMAP RFI detection, the findings of this article will be evaluated in the testbed and updated for other real-world scenarios in the future.

SMAP is designed to measure brightness temperature, operating within the protected portion of L-band, i.e., 1400–1427 MHz, to estimate soil moisture and detect the freeze or thaw state in global scale [3]. However, the SMAP measurements can be jeopardized by the corrupted RFI signals [4]. SMAP uses an on-board processing unit to gather information to identify RFI and mitigate the corrupted measurements accordingly [5]. Multiple different RFI detection algorithms are utilized, depending on which domain the interference has been eradicated such as polarization, time, frequency, code, and space. These include time domain or pulse detection, cross-frequency detection, kurtosis detection, and polarization detection [6]. RFI experienced by SMAP could be pulsed or continuous wave (CW) [7]. While pulse [8] and kurtosis [9] detection algorithms [10] are sensitive to pulsed RFI, cross-frequency detection works best on CW RFI [11]. SMAP has nine different types of detection algorithms that are combined with a logical “OR” operation to label a pixel whether it is RFI contaminated [12], [13]. All the above-stated detection schemes depend on a hypothesis that involves prior

assumptions on RFI characteristics and designed handcrafted features and algorithms specific to different RFI types.

In this article, we propose that deep learning (DL) can be an attractive alternative for RFI detection by learning directly from the data. Instead of combining multiple approaches, a single learning-based approach is demonstrated that could provide a generalized RFI detection globally. DL has become very instrumental in various classification and recognition problems in computer vision [14], [15]. Recently DL has also been utilized in RFI detection and mitigation and shown to outperform classical approaches. In [16], a DL approach is introduced to detect RFI in C-band Sentinel-1 synthetic aperture radar data. This approach utilized convolutional neural network (CNN) on RGB images generated from Sentinel-1 to detect RFI over the Tel Aviv region. Another study [17], where CNN is used in simulated time, order radio data that are acquired from a radio telescope to identify and mitigate RFI. RFI detection with DL for SMAP has been introduced in [18] where pretrained CNN models from camera images are trained and validated on 5014 spectrogram images. Spectrogram images were labeled using information from the SMAP ground processing unit and RFI quiet parts of the globe, which is validated with single cross-validation (CV) technique. Developed spectrograms were tested over Europe and the Middle East orbit utilizing only antenna powers that can be referred as the second raw moment from the horizontal polarization channel.

The proposed architecture includes both convolutional and FC layers and works on antenna counts of multiple raw moments in horizontal polarization (H-pol) and vertical polarization (V-pol) along with Stokes parameter observations directly. A supervised learning framework is developed, where utilized raw inputs from SMAP's level 1 A data are dynamically labeled using SMAP level 1B data product's quality flag eliminating the need for manual labeling. Instead of depending on different algorithms for detecting different types of RFI, the single DL framework proposed in this study has been shown to detect different types of RFI successfully on a global scale. The evaluations of the proposed approach using spatial and time-based CV approaches show that the proposed DL approach has a high level of generalization performance. The main contributions of the proposed approach can be summarized as follows.

- 1) A dedicated DL architecture with CNN and FC neural network layers is proposed that can utilize antenna counts of multiple raw moments in H-pol and V-pol along with Stokes parameters.
- 2) A step-by-step approach to create input spectrograms by taking advantage of SMAP's level 1 A data and dynamically labeling them with level 1B data products.
- 3) Over 50 million footprints are observed globally to prepare training and testing spectrograms, where the DL-based framework achieves the highest 99.99% accuracy in RFI detection with H-pol antenna counts. This helps in developing a generalized and robust RFI detection algorithm.
- 4) The proposed DL architecture is assessed under four different train/test scenarios that include spatial and time-based techniques, which help in understanding the performance of RFI detection under various scenarios.

The rest of this article is organized as follows. Details of the utilized SMAP data, dataset preparation, data statistics, and example spectrograms are discussed in Section II. Section III details the preprocessing of data, the proposed DL architecture, training of the DL architecture, and evaluation metrics of DL models. Results and discussions are provided in Section IV. Finally, the conclusion is drawn in Section V.

II. DATASET

SMAP's level 1A [19] and level 1B [20] data products are used in this study to effectively create a DL-based RFI detection framework. These are open-source data products available for scientists and researchers all over the world. SMAP's coverage started on 31 March 2015 and it is still providing valuable measurements globally. The instantaneous area of the earth that is covered by the SMAP radiometer, which is known as the footprint, is 36×47 km and SMAP takes two to three days to perform a global coverage.

A. SMAP Level 1 A Data

Level 1 A data product contains antenna counts in both full band and sub-band levels [12]. These antenna counts are provided as first-, second-, third-, and fourth-order statistical raw moments. The first raw moment (M_1) acts as the mean of the received signal. The second raw moment (M_2) is related to the variance of the signal. Consecutively, the third raw moment (M_3) gives an impression of skewness, and the fourth raw moment (M_4) is related to kurtosis. All of these moments are available in both in-phase (I) and quadrature channels (Q) of H-pol and V-pol. The j th-order raw moment is given as

$$M_j = \frac{1}{N} \sum_{i=1}^N X_i^j \quad (1)$$

where X_i is the i th raw voltage value and N is the total number of samples. Each sub-band raw moment data M_j in SMAP's level 1 A data product is stored using a four-dimensional array of size $779 \times 1928 \times 16 \times 4$, which is depicted in Fig. 1. The first dimension demonstrates the antenna scans, while the second dimension represents the number of science data packets containing 1.2 ms of information in the antenna radiometric state. The third dimension represents the 16 frequency subchannels, which cover 1.5 MHz each and together form the total SMAP radiometer band of 1400–1427 MHz [21]. The fourth dimension stores in-phase (I) and quadrature (Q) components of V-pol and H-pol channels. SMAP level 1 A data products also include third Stokes (3S) and fourth Stokes (4S) parameters, which are complex correlations between the raw moments of H-pol and V-pol signals. Details about Stokes antenna parameters are given in [22]. Stokes parameters are given in three-dimensional arrays of size $779 \times 1928 \times 16$, which are divided into antenna scan, science data packets, and sub-bands, respectively. Level 1 A data products are used to create the input dataset for the developed DL architecture.

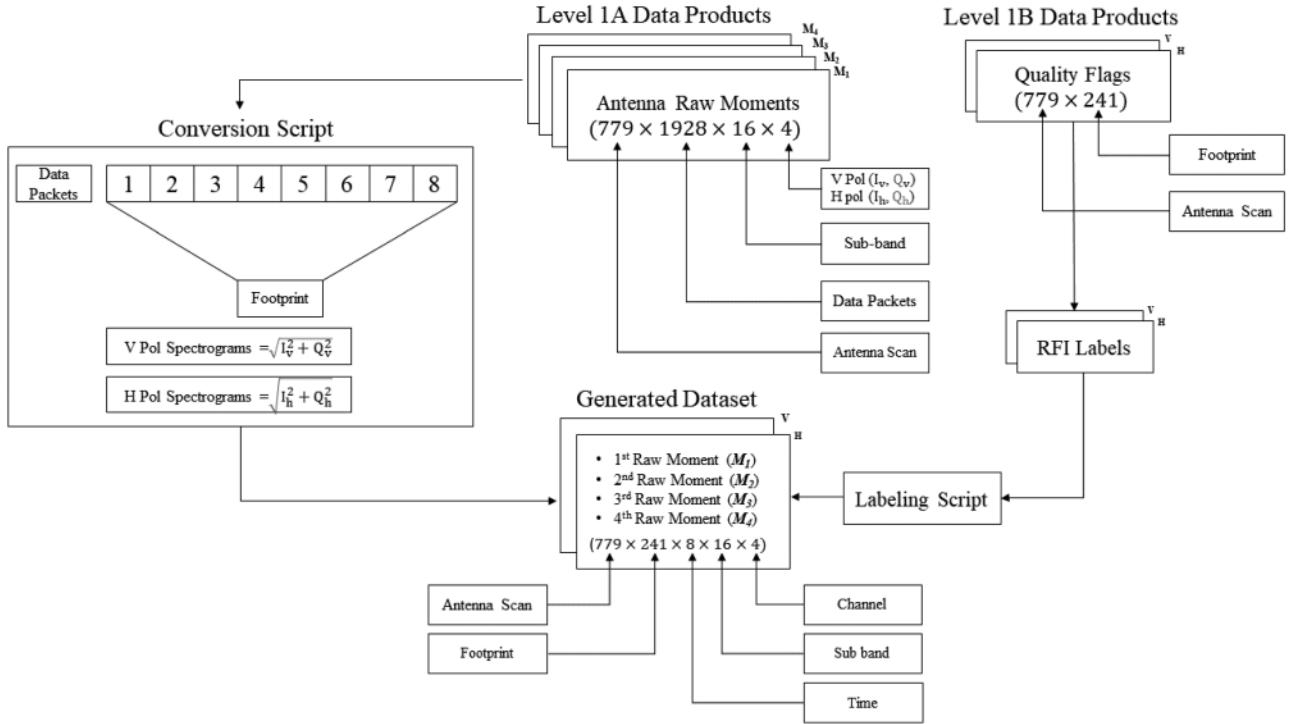


Fig. 1. Illustration of data products used for model development and labeling for RFI detection.

B. SMAP Level 1B Data

SMAP level 1B data products contain antenna temperatures, Earth brightness temperature, and quality flags [23], [24]. These flags contain 16-bit data, where each bit represents a piece of certain information. A particular bit from the 16-bit information (bit no. 3) indicates whether a certain footprint in an antenna scan is RFI contaminated or not. Level 1B quality flags are provided in categories of V-pol, H-pol, 3S, and 4S. Each category of flags is a two-dimensional array of size 779×241 , as depicted in Fig. 1, representing whether each antenna scan and footprint is exposed by RFI. Note that each category has its own individual and different quality flag data. This level 1B data product is used to label the corresponding category of level 1 A data products generating labeled datasets for supervised training and testing of the DL framework.

C. Data Preparation and Labeling

SMAP Level 1 A and 1B data are utilized to prepare the labeled datasets used in this study, as illustrated in Fig. 1. Each raw moment in a single level 1 A data file has a shape of $779 \times 1928 \times 16 \times 4$. For each footprint's sub-band, eight radiometer data packets are allotted [25]. So, in each level 1 A moment file, we have 779×241 antenna scans and footprints, i.e., $1928 = 241 \times 8$. For each antenna scan and footprint, we generate 16×8 spectrogram images of V-pol and H-pol by taking the magnitude of in-phase and quadrature of corresponding H-pol and V-pol channels. Hence, a single-level 1 A file for a particular raw moment can generate 779×241 different spectrograms for each of the H-pol and V-pol channels. First, four raw moment data are provided in SMAP level 1 A data products.

For each polarization, we combine the spectrogram images of each moment (1-4) as a new image channel creating a tensor of $16 \times 8 \times 4$ that will be input to the DL framework, as detailed in the next section. This process generates separate moment spectrogram tensors for each polarization as well as for each antenna scan and footprint. Using the Level 1B data flags, we label each spectrogram tensor as RFI contaminated or not. Third and fourth Stokes parameters are converted into spectrograms similarly, creating single channel 16×8 size images. They are labeled using their corresponding quality flag level 1B data. Both level 1 A and level 1B data products are coherent in terms of time, antenna scan, and footprints. By utilizing this homogeneity, each constructed input type is dynamically labeled with the SMAP quality flags.

This process generates a total of four different datasets: four-channel moment spectrogram tensors for H-pol and V-pol, and single-channel spectrograms for third Stokes and fourth Stokes parameters. Since RFI labels for different types of data are provided differently by SMAP, each dataset has its own individual and different RFI label. A particular data file from SMAP can generate as many as 187 739 footprints, i.e., $187,739 = 779 \times 241$, and multiple files are utilized to gather samples related to RFI-contaminated and RFI-free footprints, which is discussed in the following section. SMAP quality flags are considered as the ground truth for this study, which might have its own false alarm rate. Moreover, there can be instances of missed detection with SMAP algorithms [26]. Considering the challenges associated with verifying these large datasets, this study utilizes SMAP level 1 A antenna counts and level 1B quality flags as training and testing samples.

TABLE I
DATA STATISTICS OF RFI AND NO-RFI FOOTPRINTS

Year	Antenna Counts Domain	Observed RFI Footprints	Observed No-RFI Footprints	RFI:No-RFI (%)	Common RFI	Common No-RFI
2017	Vertical	189,562	19,687,569	0.9629	20,651	19,534,813
	Horizontal	198,285	19,651,326	1.0090		
	Third Stokes	70,379	19,813,780	0.3552		
	Fourth Stokes	65,064	19,818,515	0.3283		
2018	Vertical	163,471	19,627,873	0.8329	16,681	19,426,877
	Horizontal	166,335	19,531,635	0.8516		
	Third Stokes	57,371	19,740,307	0.2906		
	Fourth Stokes	51,910	19,745,250	0.2629		
2019	Vertical	160,144	19,675,312	0.8139	15,107	19,435,689
	Horizontal	164,453	19,537,851	0.8471		
	Third Stokes	54,813	19,785,713	0.2770		
	Fourth Stokes	50,147	19,789,918	0.2533		
Total	Vertical	513,177	58,990,754	0.8699	52,439	58,397,379
	Horizontal	529,073	58,720,812	0.9010		
	Third Stokes	182,563	59,339,800	0.3077		
	Fourth Stokes	167,121	59,353,683	0.2816		

D. Data Statistics

To construct spectrograms of RFI and No-RFI features, SMAP's level 1 A and level 1B data products are used. These spectrograms are utilized to detect RFI-contaminated footprints within a DL framework. As described earlier, four raw moments of V-pol and H-pol along with Stokes parameters are collected as training and testing datasets. Samples used in this study are collected from June 1 to June 4, 2017, September 1 to September 4, 2018, and March 1 to March 4, 2019, which took a total memory of over 600 GB in the computer hardware. SMAP usually takes 2–3 days to perform a global coverage and data are taken in a way so that there are instances of RFI and No-RFI cases across the globe. Moreover, samples are accumulated from different years to understand whether RFI types change over time along with the effectiveness of the detection algorithm. All the data files from the mentioned time frame from SMAP level 1 A and level 1B are utilized to prepare and label the spectrograms.

Table I lists the number of SMAP footprints inspected for the observed time spans each year for varying antenna count domains. For each antenna count domain such as V-pol, H-pol, 3S, and 4S, the number of RFI observed footprints and RFI free footprints is given along with the total numbers over the full-time span tested.

For a fixed time frame and antenna count domain, it can be seen that the number of footprints flagged as RFI by horizontal and vertical quality flags is comparably higher than third and fourth Stokes quality flags. The number of RFI-flagged footprints for each domain has slightly decreased each year. From the observations, it can also be seen that the ratio of RFI label footprints to the number of no-RFI footprints is very small, in most cases below 1%. This leads to a highly imbalanced dataset where comparably a very small number of RFI cases are detected compared to no-RFI cases. It is important to use proper metrics to understand the performance of the detection algorithm in this highly imbalanced dataset, which is detailed in the Section III-D. Different numbers of RFI footprints for varying antenna count domains indicate that a single footprint may be flagged as RFI from one domain while being flagged as no-RFI in another domain. From the observed samples, common RFI footprints

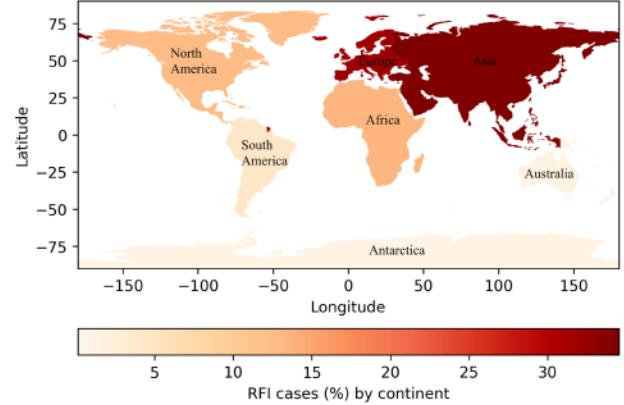


Fig. 2. Distribution of RFI cases by continent.

are also identified, where a certain footprint is labeled as RFI by all four different quality flags. Comparably smaller number of common RFI footprints indicate low agreement between SMAP flags for each domain.

The spatial distribution of RFI cases across the globe is portrayed in Fig. 2. The number of RFI-contaminated footprints is segmented into seven continents. Total common RFI cases are demonstrated in Table I and are used to identify the continents with a higher number of RFI cases. Other quality flags show a similar spatial distribution of RFI-contaminated footprints. Among the continents, Asia and Europe are responsible for the majority of the RFI cases globally with 34.5% and 31.9%, respectively. The performance of the proposed DL framework is evaluated in different regions of the world and region-based analysis, showing high generalization performance for the overall detection algorithm, and is provided in Section IV.

E. Example Spectrograms

In Fig. 3, we show spectrograms for three example footprints categorized as "Common RFI," "Common No-RFI," and "Mixed Cases." As described in previous sections, spectrograms are divided into four different categories as V-pol, H-pol, 3S, and 4S. Moreover, V-pol and H-pol datasets consist of four different

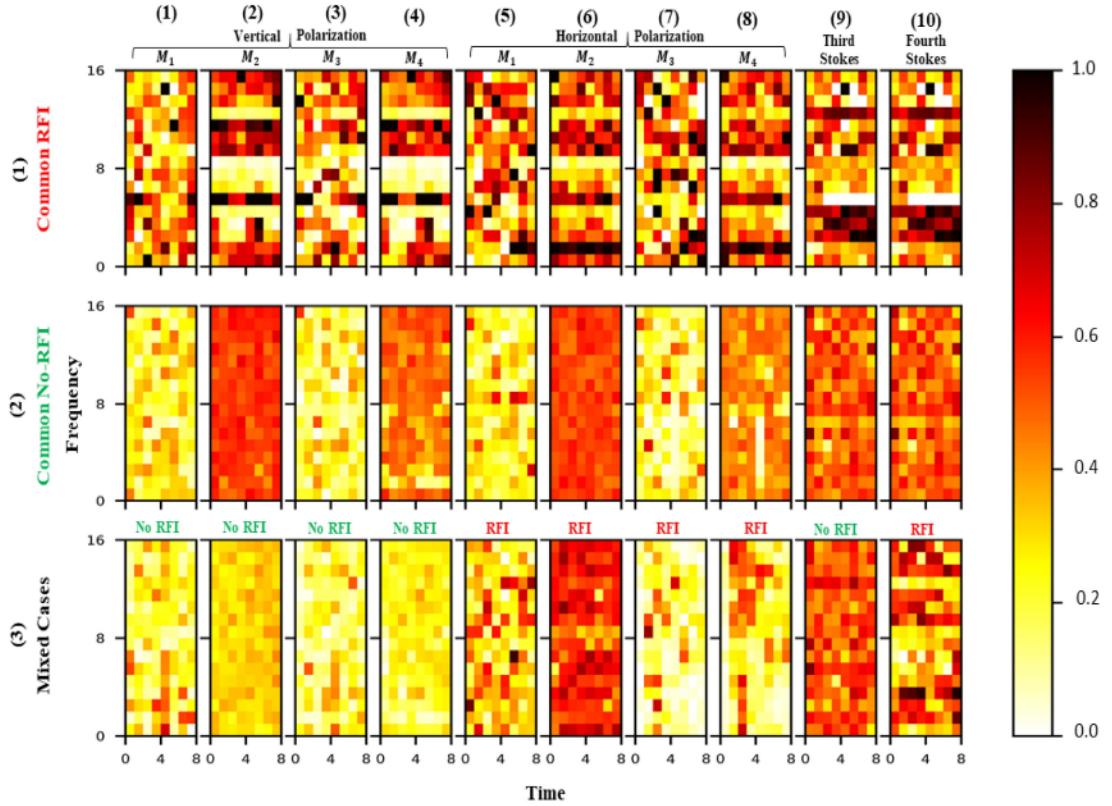


Fig. 3. Example spectrograms generated from level 1 A data and labeled by label 1B data. Columns (1–4) and (5–8) show V-pol and H-pol spectrograms, respectively. Columns (9) and (10) exhibit spectrograms with 3S and 4S. Row (1) illustrates a particular footprint which is labeled as RFI by all four quality flags. Row (2) demonstrates common No-RFI cases. Row (3) indicates a particular footprint which is labeled differently (RFI by H-pol and 4S but No-RFI by V-pol and 3S) by four different quality flags.

statistical raw moments. Antenna counts of generated spectrograms are normalized between 0 and 1, which is detailed in Section III-A. Thus, each row of Fig. 3 shows a total of ten spectrograms from all four categories for a single footprint. Columns (1–4) and (5–8) demonstrate the V-pol and H-pol spectrograms of four different raw moments respectively. Columns (9) and (10) show spectrograms of the third and fourth Stokes parameters. The spectrograms in row (1) are generated from a particular footprint which is labeled as RFI by all four quality flags. The second footprint spectrograms shown in row (2) are from an example footprint where every quality flag specifies that there is no-RFI. This case is depicted in the figure as a “common no-RFI.” In “common RFI” and “common no-RFI” cases, there are visible differences in spectrogram images that can be indicative of the RFI. The third example footprint is a mixed case where it is labeled as “RFI” by H-pol and 4S quality flags but “No-RFI” by V-pol and 3S quality Flags. It is difficult for SMAP algorithms to detect moderate (10–100 K) and lower level RFI (5–10 K) [13] and these examples might contain information about low and moderate level RFI. All these spectrograms are considered for training and testing in the proposed DL architecture.

III. METHODOLOGY

DL is one of the subsets of machine learning, which involves a parametric model to learn from various types of data, such

as images, videos, or speech [15]. CNN is one of the highly used DL network due to its high feature learning capability and it plays an important part in modern-day computer vision. This section details the structure and training of a DL architecture designed specifically for the RFI detection from SMAP data. In DL, there are various existing large-scale networks trained on millions of camera images for object classification purposes, such as AlexNet [27], ResNet [28], and VGG Net [29]. Although these pretrained networks could be used within the transfer learning [30] framework for RFI detection as in [18], the characteristics of spectrograms are very much different than camera images. While transfer learning is more suitable for limited data to utilize the feature extraction capability of a pretrained network, the generated large dataset and labels in this study allow utilization and learning of a new model for RFI detection. Evaluation and performance metrics are discussed in the latter parts of this section where different train/test scenarios are explained with appropriate performance metrics to explain the detection performance in an imbalanced dataset.

A. Data Preprocessing

Four different datasets are generated, as described in Section II, to evaluate the DL framework. The spectrograms are prepared with the magnitude of in-phase and quadrature values. In a previous DL-based RFI detection work, similar spectrograms are converted into RGB images followed by a

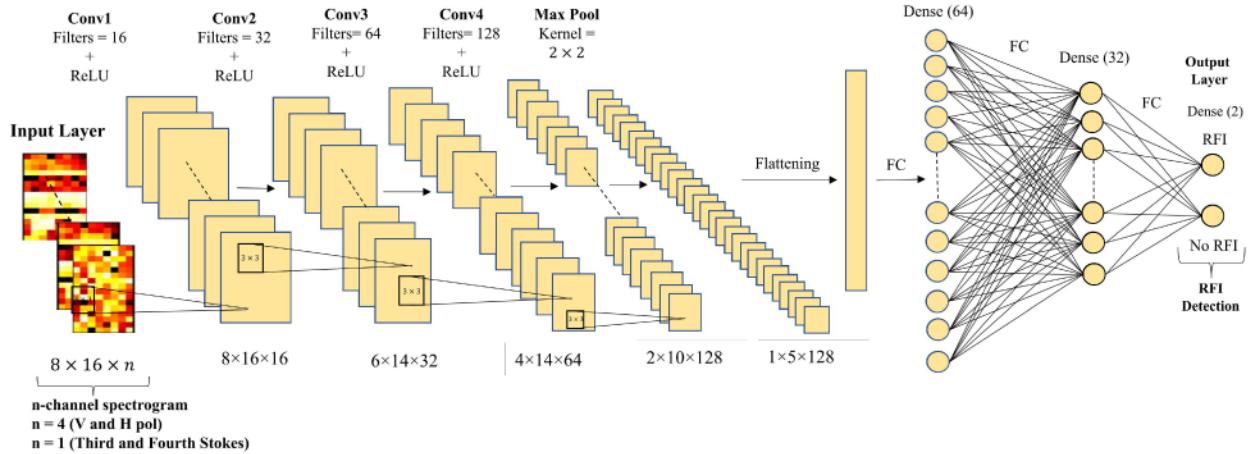


Fig. 4. DL architecture to detect RFI.

min-max normalization to feed the data into pretrained networks on camera images [18]. Converting spectrogram to RGB images has the potential to limit the range of image intensity into integers from 0 to 255 affecting the dynamic range inherent in the data. Antenna counts of different polarization and Stokes parameters can have very high numerical values because of incorporating higher order moment data [21]. Turning spectrograms of antenna counts to RGB images will add an additional preprocessing layer which might be responsible to loose important information about RFI. In this study, before being sent into the DL framework to detect RFI, spectrograms are directly normalized using the min-max normalization technique rather than being converted into RGB images.

In the min-max normalization technique, the minimum value of a spectrogram gets transformed into a 0, the maximum value gets transformed into a 1, and every other value gets transformed into a decimal between 0 and 1. Normalization is performed on the antenna counts of the whole dataset consisting samples from land and ocean before training and testing with the model. As different polarization datasets are trained and tested separately, samples are normalized for each dataset independently. Normalization is a preprocessing approach that helps in converging the model with less computation complexity [31]. This allows the DL architecture to learn directly from the spectrogram tensors by keeping the existing dynamic range of data. Other normalization techniques, such as utilizing mean and standard deviation, are also inspected in this study but min-max normalization provides the best result in terms of accuracy and computational time.

B. Design of the DL Architecture

A DL architecture is proposed formulated with convolutional and fully connected (FC) layers that will map the input raw moment spectrograms to a binary classification model where No-RFI cases are considered as “0” (negative) and RFI cases as “1” (positive). The developed architecture has been illustrated in Fig. 4 for the spectrogram tensor inputs. The proposed architecture has four convolutional layers followed by two FC layers and a final dense layer with two neurons accompanied

by a softmax activation function to represent RFI and no-RFI classes. The goal of the convolutional layers is to learn features directly from the input spectrograms and the FC layers map these features to the output RFI detection. A binary cross-entropy loss function is associated with the DL architecture that updates the learning parameters by comparing them with the ground truth after each iteration. The output after the final soft-max layer can be interpreted as probabilities so that the architecture as a whole maps input spectrograms to probabilities of whether that input has an RFI or not.

The input to the DL architecture is $16 \times 8 \times 4$ moment spectrogram tensor for V-pol or H-pol cases and 16×8 matrices for 3S and 4S parameters. Other than the first layer that accepts these inputs, we used the same DL architecture for all input cases. The first convolutional layer started with 16 filters, which have 3×3 kernels. Then with each CNN layer, filters are increased to 32, 64, and 128 and all have 3×3 kernels. While the first layer used the same padding, no zero padding is used for the following layers. After convolutional layers, a two-dimensional max-pool layer is used to extract low-level features of the inputs. The extracted features are flattened and input to the FC layers with 64 and 32 neurons, respectively. After each layer, a rectified linear unit activation function is used [32]. This activation function helps with the vanishing gradient and saturation problems [33]. A soft-max layer follows the final layer, which outputs the probabilities of RFI and no-RFI classes. The whole DL architecture is built using TensorFlow Keras API [34].

C. Training the DL Architecture

The flowchart of the training process for the proposed DL architecture is depicted in Fig. 5. We trained and tested the detection model with four different datasets: V-pol, H-pol, 3S, and 4S parameters. Each model maps the input spectrograms to a final output of RFI or no-RFI probabilities. The only difference in training and testing with four datasets is the input layer and corresponding labels for each dataset, which are detailed in the previous section. The model is trained and tested by maintaining

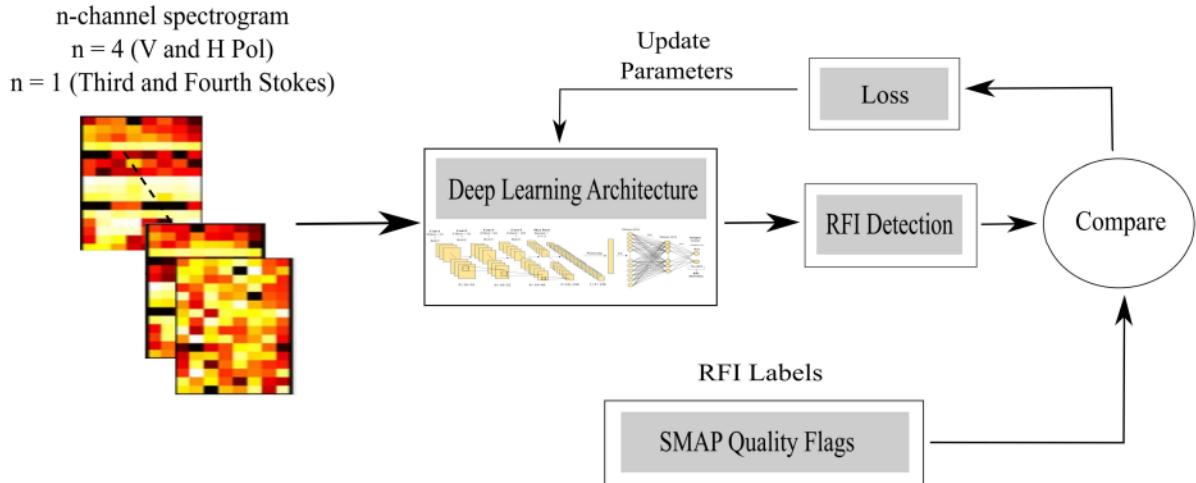


Fig. 5. Flowchart of the proposed RFI detection algorithm using spectrograms.

the ratio of “RFI” and “no RFI” cases discussed in Section II-D to emulate a real-world scenario.

To update the parameters of the DL architecture, the outputs of the model are compared with the RFI labels, which are SMAP quality flags for the corresponding inputs. The model parameters are determined in order to minimize the binary cross-entropy loss function [35]. The minimization of the loss function is achieved through a version of the gradient descent-based backpropagation approach. Adam optimizer is used for this purpose, which iteratively updates the model parameters, and helps with computational speed [36]. To develop a model that generalizes and does not overfit to the training data, learning rate schedulers [37] and early stopping [38], [39] have been introduced to the model. The learning rate is a hyperparameter that governs how much the model changes each time the model weights are changed in response to the predicted error. A smaller value may result in a long training process, whereas a larger value may result in an unstable training operation that might not converge. Learning rate schedulers refer to the idea of decreasing the learning rate after each iteration of the model and this study utilizes an exponential learning rate scheduler. It helps to decay the learning rate exponentially with an initial learning rate of 0.01 and gradually decreased the learning rate after each epoch. Early stopping checks the accuracy and losses in between the iterations to force stop the training process before it starts overfitting.

A total of 20 epochs are used with a minibatch size of 10,000 data samples for training based on the convergence pattern of the model. Different training/testing scenarios that lead to various validation techniques are used to evaluate the model, which is depicted in Section III-D. Python is used to develop conversion scripts discussed in Section II-C and the DL framework.

D. Evaluation and Performance Metrics

CV is an important process to evaluate the generalization and effectiveness of a DL-based model. It is most commonly utilized in situations when the goal is prediction or classification and how well a model will perform in practice. Four types of CV

have been implemented in this study: train-test split, fivefold, time-based, and region-based. To author’s knowledge, no other study in DL-based RFI detection has evaluated their algorithm with four different CV techniques. Each CV technique is trained and tested with the four different datasets of V-pol, H-pol, 3S, and 4S spectrograms.

At first, the train-test split technique is deployed, where total data are split into 80% for training and 20% for testing/validation randomly for the whole world and time span. Second, K -fold CV has been used, where data products are randomly split into K folds ($K = 5$ for this study) for training and testing. While the DL is trained on $K - 1$ fold data, it is tested on the remaining fold and average performance results are reported on overall K folds. Train-test split and K -fold are two conventional approaches in DL to evaluate a model’s performance [40].

After evaluating traditional CV techniques, training and testing datasets are divided with respect to different time spans. The main goal of this type of evaluation method is to understand whether the characteristics of RFI change over time and if it can be modeled with data products from a certain period of time. Finally, the dataset is divided regionwise, where the DL model is trained with samples from different regions around the world and tested with regions that are not considered in training. Region-based CV portrays an important analysis of whether RFI detection models can be trained in a certain region and successfully used for other regions. This also helps to understand the spatial distributions of different types of RFI that can be generated from diverse sources.

The confusion matrix is generated to define the performance of RFI detection and it helps to visualize and summarize the overall capability of the detection framework. From the confusion matrix, other performance-indicating metrics, such as accuracy, precision, recall, and F1 scores, are generated. An example confusion matrix is given in Table II for the binary classification problem with true and predicted classes of RFI and No-RFI. Depending on the predictions the elements of the confusion matrix will be as shown, where TP = True Positive, TN = True Negative, FP = False Positive, and FN = False

TABLE II
GENERAL CONFUSION MATRIX FOR DETECTION

True Class	Confusion Matrix	
RFI	TP	FN
No-RFI	FP	TN
	RFI	No RFI
	Predicted Class	

Highlighted portion of the table portrays the significance of true detection.

Negative. “RFI” cases are considered as “1” or positive class and “No-RFI” cases are considered as “0” or negative class. Using the confusion matrix, the performance metrics are calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}}. \quad (5)$$

Accuracy is perhaps the highly used metric to understand the overall classification performance of the model. However, a high accuracy does not always indicate the general detection performance of the method, especially on an imbalanced dataset scenario as in this case, where the number of “No-RFI” cases is very larger than the “RFI cases.” Metrics such as precision and recall should also be evaluated to observe the general performance and they help to understand the performance with a large biased dataset in terms of false positives and false negatives, which is very critical in RFI detection [41], [42]. Precision refers to how precise/accurate a model is in terms of how many of the anticipated positives are actually positive. Precision is a good statistic to employ when the costs of false positives are high. On the other hand, recall determines out of all actually positive cases (in this case “RFI”), how many the model predicted to be positive. When there is a large cost associated with a False Negative, recall can be the measurement metric that will help to select the optimal model. In this study, a higher precision and recall value can be important metrics to determine the performance of RFI detection. F1-score is a function of both precision and recall creating a single metric instead of two and is high when both precision and recall are high. F1-score can also be referred to as the overall performance of a classification model. All the stated performance metrics are evaluated in the testing dataset, which has not been seen by the model during training.

Receiving operating characteristics (ROC) is another illustration to define the overall RFI detection performance of a model [43]. ROC basically shows the probability of detection (true positive rate) as a function of the false alarm rate (fraction

TABLE III
CONFUSION MATRIX (%) OF RFI DETECTION

Antenna Counts Domain	True Class	Train-Test Split		K-Fold	
		RFI	0.298	99.992	0.008
Vertical	No-RFI	0.001	99.999	0.001	99.999
	RFI	99.974	0.026	99.993	0.007
Horizontal	No-RFI	0.001	99.999	0.002	99.998
	RFI	99.66	0.34	99.997	0.003
Third Stokes	No-RFI	0.12	99.88	0.252	99.748
	RFI	99.74	0.26	99.902	0.098
Fourth Stokes	No-RFI	0.03	99.97	0.036	99.964
	RFI	No RFI	RFI	No RFI	
		Predicted Class			

Highlighted portion of the table portrays the significance of true detection.

of false positive). Evaluating ROC curves for machine learning-based classifications are detailed in [44]. The normalized area under the ROC curve (AUC) is being used as a performance indicator to estimate the relative performance of detection algorithms under diverse scenarios. A higher AUC means a higher performing detection algorithm.

IV. RESULTS AND DISCUSSION

In this section, analysis of the RFI detection performance of the proposed DL model has been demonstrated. Spectrogram tensors of V-pol, H-pol, 3S, and 4S detailed in Section II, containing the information of antenna counts with labels are utilized for training and testing. The DL framework has been trained and evaluated under various scenarios using performance metrics described in Section III-D. Overall performance in recognizing RFI and No-RFI scenarios with train/test split and K -fold is depicted in Section IV-A. The dependence of the dataset is explained in Section IV-B. Day-based and region-based CV analysis is illustrated in Sections IV-C and IV-D, respectively.

A. Overall Performance of DL-Based RFI Detection

In this section, the performance of this study’s DL-based RFI detection algorithm is shown with four different datasets and two CV techniques of train/test split and K -fold. Table III lists the confusion matrix for all antenna count domains of V-pol, H-pol, 3S, and 4S under both CV techniques. The performance metrics of RFI detection, such as accuracy, precision, recall, and F1-score, are computed from the confusion matrix and given in Table IV for all tested scenarios. Utilizing and validating four different datasets with a single coherent DL model portray the flexibility of DL-based model in RFI detection. Maintaining the ratio of RFI and No-RFI cases given in Table I, 80% of the total samples for each antenna count domain are used for training and 20% for testing in the train-test CV technique. From this analysis, accuracy with all four different datasets reaches over 99.7% and best achieved at 99.99% for H-pol. Because of the class imbalance associated with “RFI” and “No-RFI” samples, accuracy alone might not be an ideal metric in these scenarios to evaluate the success of a detection algorithm. Precision of RFI

TABLE IV
PERFORMANCE METRICS (%) OF RFI DETECTION

Validation Techniques	Antenna Counts Domain	Accuracy	Precision	Recall	F1-Score
Train-Test Split	Vertical	99.85	99.99	99.70	99.85
	Horizontal	99.99	99.89	99.97	99.93
	Third Stokes	99.77	99.88	99.66	99.77
	Fourth Stokes	99.86	99.97	99.74	99.86
K-Fold	Vertical	99.99	99.99	99.99	99.99
	Horizontal	99.99	99.99	99.99	99.99
	Third Stokes	99.87	99.75	99.99	99.87
	Fourth Stokes	99.93	99.96	99.90	99.93

detection is also above 99.7% for four different datasets, which prove that the algorithm detects RFI cases successfully with very few false positive associated with it. The highest precision level in train/test split CV is achieved for V-pol case with 99.99%. This study's DL-based algorithm also produces a high recall score in all datasets indicating a very low false negative rate in overall RFI detection. F1-score is a combination of precision and recall, which is greater than 99.8% for each dataset portraying the overall performance. Comparing all the datasets in the train-test split, H-pol and V-pol provide the best overall metrics in detecting RFI.

K-fold shows similar results where the overall dataset is divided into $K = 5$ different folds and then trained in $(K-1)$ fold and tested in the remaining fold repeating the process for each fold independently. Results generated in each fold are averaged to obtain the results, which are also given in Tables III and IV. It can be seen that specifically for V-pol and H-pol cases, each performance metric is 99.99%, which is an indication that a very high rate of tested samples is correctly classified. From the analysis, all four datasets provide satisfactory and similar performance in RFI detection but V-pol and H-pol cases produce slightly better precision and recall than Stokes parameters. This shows that the proposed DL framework learns the SMAP RFI flagging for V-pol and H-pol very accurately.

During the training of a DL-based model, it is important to understand whether the model is overfitting or underfitting. Overfitting means when a model works well in a training scenario but cannot generalize in testing samples and a model is considered as underfitting when it underperforms in both training and testing data. It is very important to find the optimum spot in DL, which is considered as a good fit. When train and test/validation loss is decreasing after each epoch and converges after some number of iterations together at a low loss level, it is considered as a good fit in DL, as depicted in Fig. 6. Both training and validation loss are decreasing with each epoch, which demonstrates that the model is neither underfitting nor overfitting. The model loss here is plotted for the H-pol dataset and other datasets show similar results.

These models have been trained and tested on a machine with Intel(R) Xeon(R) 4116 CPU, 128 GB memory and NVIDIA TITAN RTX GPU. The total training time for the fivefold CV over the whole dataset is 5.76 h and the testing time is 0.73 h. On average, predicting whether a particular footprint's spectrogram has RFI or not takes 0.13 s for the proposed DL approach. Noting that the implementation is not being optimized for computation time, this duration can be potentially reduced. This might be an

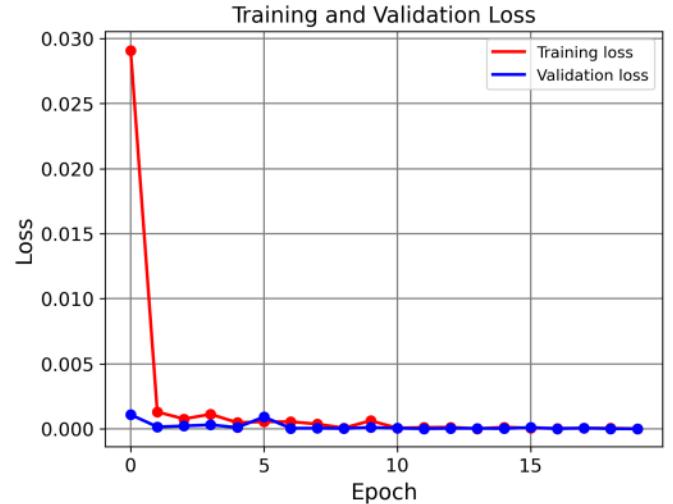


Fig. 6. Model loss on training and validation dataset on H-pol dataset.

important consideration for future space-borne missions to include an onboard RFI detection unit running DL-based solutions in real time.

A comparison of AUC with the help of ROC curves between all the datasets with the train-test CV technique is depicted in Fig. 7. The ROC of any detection algorithm aids in the visual representation of the probability of detection (true positive rate) versus the false alarm rate (false positive rate). A higher AUC helps in determining the relative performance of the detection algorithm. Traditional approaches used for RFI detection in passive remote sensing are also compared with this study's DL model evaluated on all test datasets. SMAP has nine different detection algorithms applied in V-pol and H-pol antenna counts, which are combined with an "OR" operation to increase the performance of detection [13]. But this study utilizes a single DL-based architecture to detect RFI globally. Among the traditional approaches, pulse and kurtosis detection algorithms are evaluated. Pulse detection compares the deviation of a particular measurement and kurtosis utilizes the ratio of the second and fourth raw moments. Details about these algorithms can be found in [45]. ROC for both of these algorithms is directly taken from the SMAP's algorithm theoretical basis document [12]. A positive sloped ROC curve that spans diagonally across the figure area depicts a detector with a "50/50" guess as to whether RFI is present. ROC curves are demonstrated with both linear [see Fig. 7(a)] and logarithmic [see Fig. 7(b)] scales specifically to illustrate the DL-based ROC performance curves, which provide high detection rate with very small false-alarm rates. The highest AUC achieved through the DL framework is approximately 0.9999 with H-pol and V-pol datasets. DL-based RFI detection performance is significantly higher than the traditional RFI detection techniques of kurtosis, which provide AUC of 0.85 and the pulse detection algorithm provides AUC of 0.64. AUC with 3S (0.9990) and 4S (0.9995) parameters also show better performance than traditional approaches suggesting DL provides better overall performance in all four data products. To illustrate ROC curves and calculate AUC, Python's "Scikit-Learn" package is used in this study [46].

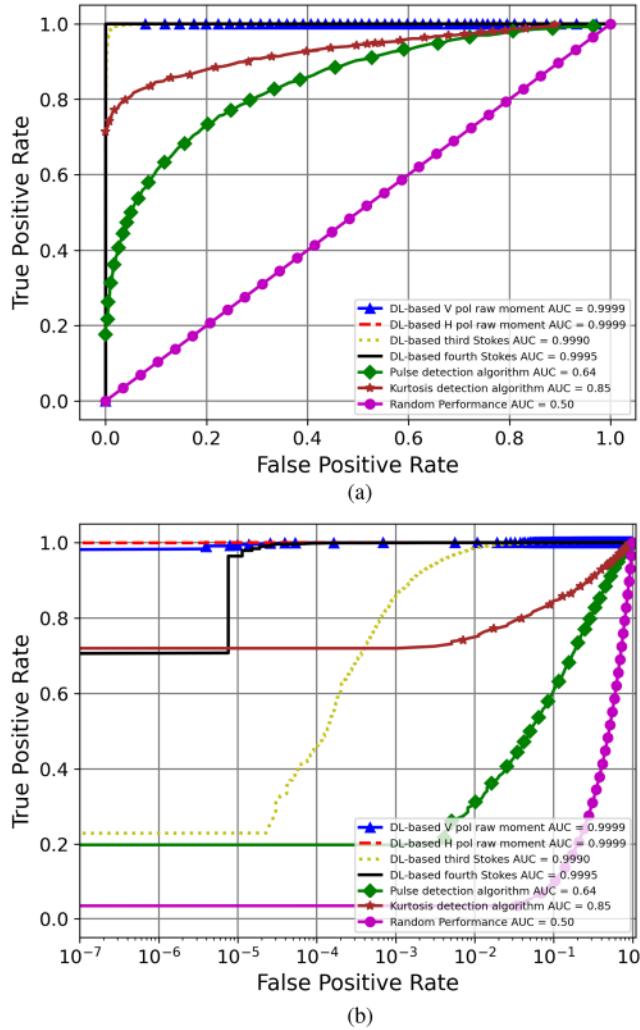


Fig. 7. Performance analysis with ROC curves (a) DL-based performance with four different dataset and traditional approaches (b) with logarithmic scale.

B. Data Dependence

A data-driven technique, such as DL, is designed to learn and make decisions directly from the data. In this study, SMAP's flags are used to label the antenna count data but experts labeling the same data, collecting samples all around the world over the desired time span can be very difficult. Thus, it is crucial to know the effect of the number of training samples required to train a model for satisfactory RFI detection. In the previous section, the samples collected globally are divided into 80% training and 20% testing to evaluate the performance for each antenna count domain. In this experiment, the test dataset is fixed and to train the DL-based detection model, different rates of samples are used. H-pol antenna counts are used in this experiment where train and test samples are randomly taken from the total observed footprints maintaining the ratio of RFI and No-RFI cases. The achieved performance metrics as a function of training rate are illustrated in Fig. 8. We started the experiment with 5% of total training samples and gradually increased the sample numbers with the highest being 80%. The performance metrics utilized in this study such as precision, recall, and F1-score increase with

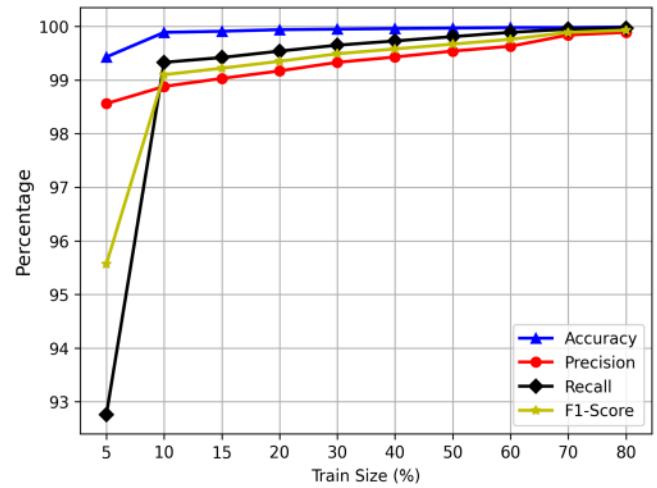


Fig. 8. Performance metrics as a function of training data size in DL framework for RFI detection.

TABLE V
PERFORMANCE METRICS (%) FOR SINGLE DIFFERENT RAW MOMENTS

Antenna Counts Domain	Moment	Accuracy	Precision	Recall	F1-Score
Horizontal	M_1	99.82	93.44	76.2	83.95
	M_2	99.98	99.93	99.80	99.87
	M_3	99.96	99.87	99.59	99.73
	M_4	99.98	95.84	99.86	99.85
Vertical	M_1	98.29	91.96	83.62	87.60
	M_2	99.97	99.74	99.86	99.80
	M_3	99.93	99.86	99.12	99.49
	M_4	99.98	99.90	99.81	99.86

the training sample size. Lower number of training samples such as 15%–20% demonstrate precision and recall over 99%. This analysis shows that a DL-based algorithm can be effective with a lower percentage of training samples to detect RFI. Because of the availability of SMAP data products, we trained our model with all available training samples that help to increase the performance metrics. Moreover, a higher number of training samples help to extract features from different types of RFI conditions globally which contributes to reducing missed RFI detection.

In the previous section, the experiment is conducted with four raw moments jointly being utilized as an input to the DL model. In this section, we analyze the individual performance of these raw moment spectrograms for the overall detection algorithm. Table V lists the performance metrics obtained from individual raw moments from V-pol and H-pol datasets. In these experiments, each raw moment spectrogram is utilized as a single channel image. Utilizing only the first moment M_1 for both V-pol and H-pol provides significantly low precision and recall compared to jointly using all four channel spectrograms (demonstrated in Table IV). Recall for M_1 with V-pol and H-pol is 76.2% and 83.62%, respectively. This means that the DL model misses a lot of RFI-contaminated footprints if only trained on the first raw moment. Performance gets better with higher order raw moments such as M_2 , M_3 , and M_4 , but all individual raw moment performance is lower than joint utilization of all of

TABLE VI
PERFORMANCE METRICS (%) OF TIME-BASED ANALYSIS (TRAINED WITH 2017 AND 2018 SAMPLES)

Antenna Counts Domain	Accuracy	Precision	Recall	F1-Score
Vertical	99.80	96.85	99.99	98.4
Horizontal	99.92	98.71	99.99	99.35
Third Stokes	99.60	98.75	83.10	90.25
Fourth Stokes	99.65	95.85	85.05	90.12

TABLE VII
PERFORMANCE METRICS (%) OF TIME-BASED ANALYSIS (TRAINED WITH 2017 SAMPLES)

Antenna Counts Domain	Accuracy	Precision	Recall	F1-Score
Vertical	98.88	84.67	99.10	91.32
Horizontal	98.93	85.39	99.69	91.99
Third Stokes	97.48	82.56	81.73	82.14
Fourth Stokes	97.35	82.17	80.55	81.35

them together. For individual performances, second (M_2) and fourth (M_4) moments demonstrate better overall performance (F1-score) for both V-pol and H-pol cases. These results are encouraging and a single-channel spectrogram can also be utilized for the RFI detection algorithm. However, as satellite data products are critical for taking important decisions, four-channel spectrograms are preferable in detecting RFI and provide superior performance than a single-channel spectrogram.

C. Time-Based Analysis

In time-based analysis, training and testing samples are divided into different time frames. We have implemented two validation models. First, the DL model is trained with samples from 2017 and 2018 and tested with samples from 2019. For the second analysis, the model is trained only with 2017 samples and tested again on samples from 2019. These analyses will help to understand whether training with samples from a particular time frame is enough to detect RFI in a totally unseen time frame. The results of the first time-based CV analysis are given in Table VI, where all performance metrics are computed for the four different datasets. Both V-pol and H-pol datasets show satisfactory accuracy (99.80% and 99.92%) and recall (99.99% for both cases), but the precision is higher with H-pol indicating a lower number of false positive cases. Recall of 3S and 4S parameters are 83.10% and 85.05%, respectively, which are significantly lower than the other two datasets. Differences with performances in 3S and 4S parameters can be due to the fact that H-pol and V-pol parameters consist of higher order moments which are accommodated to the model with four-channel configuration, whereas 3S and 4S have single channel inputs. As DL has more numerical features to train the model with H-pol and V-pol parameters, it likely provides better performance in this experiment.

In the second CV analysis, the time difference between training and testing datasets is longer, where training is done in 2017 but the trained model is tested with the 2019 data. Obtained performance metrics are listed in Table VII. It can be seen that while accuracy and recall metrics are slightly reduced, the precision of

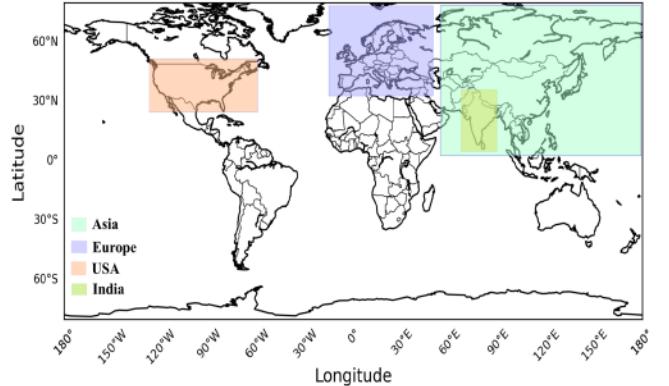


Fig. 9. Region-based analysis evaluation regions are shown with bounding boxes.

the model is significantly affected. This could be because during the long lapse between training and testing time-frames, RFI characteristics might have changed or the receiver parameters have drifted. A very important point for future research is the implementation of additional preprocessing and calibrations on raw antenna count data before they are used as input to DL models to remove any receiver-related biases. Comparing the overall performance and time-based analysis, it can be referred that V-pol and H-pol inputs can be more robust in terms of RFI characteristic changes in time and space. These findings may further indicate worldwide RFI types or characteristics have remained similar over shorter time lapses between training and testing.

D. Region-Based Analysis

For this experiment, we plan to test the effectiveness of the DL model trained on spatially different regions of the world and tested on data from a new region. To achieve this, the samples of H-pol are divided into different spatial regions to train and test the DL-based RFI detection model. Several regions are established with bounding boxes across the world, as demonstrated in Fig. 9. SMAP data products come with a specific latitude and longitude for a particular footprint, which helped to establish these regions and generate dataset corresponding to them. This analysis will help to understand the different RFI types across the globe along with the importance and effectiveness of training samples from a particular region. Train-test regions along with overall performance are given in Table VIII.

In the first experiment, the RFI detection model is tested with samples from the Europe region and trained with samples from the rest of the world (ROW). This analysis gives a perfect performance in RFI detection across Europe. When the model is trained in Europe and tested on ROW, the results deteriorate providing 97%–98% performance metrics. This analysis shows that samples from ROW have enough statistical examples to represent RFI cases in Europe while the inverse is not completely correct. Europe region might not have enough features or all possible RFI types to model global RFI cases. Similar results are also observed when trained and tested with Asia and the ROW.

TABLE VIII
PERFORMANCE METRICS (%) OF REGION-BASED ANALYSIS

Train Region	Test Region	Accuracy	Precision	Recall	F1-Score
ROW Europe	Europe	100	100	100	100
	ROW	98.49	98.13	97.12	97.625
ROW Asia	Asia	99.99	99.93	99.86	99.90
	ROW	97.26	86.37	99.97	92.68
Europe Asia	Asia	99.94	99.99	99.90	99.95
	Europe	98.27	99.99	86.78	92.92
USA USA	Europe	80	39.98	99.89	57.10
	Asia	78.17	21.68	80.27	34.14
USA	India	98.92	99.46	86.01	92.25
Europe	USA	99.98	100	99.96	99.98
Asia	USA	99.25	100	97.97	98.97
India	USA	99.34	98.2	99.75	98.94

Next, train/test data are split between Asia and Europe. These two regions possess most of the RFI cases globally and this analysis will help to understand how effectively DL-model learns RFI types and characteristics in different continents. When trained in Europe and tested in Asia, both precision and recall are very high. This means training samples from Europe contain enough features of RFI and No-RFI footprints which are successfully learned by the detection model. When the model is trained in Asia and tested in Europe, recall is comparably lower. This means the detection framework missed some RFI cases that leads to false negative decisions. This shows that samples from Asia do not fully contain the types of RFI cases that are happening over Europe. This analysis can be an indication that samples from a highly RFI-contaminated region can be utilized to establish a successful DL-based RFI detection model for other regions.

Samples collected from a continent might also be a challenge itself in terms of cost and feasibility. So, this study's DL-based model is trained with samples from a particular region (USA region for this experiment) and then tested in diverse spatially distributed regions such as Europe, Asia, and India. Testing with Europe shows that despite a high recall (99.89%), precision is around 39.98%, which is evidence of a very high false alarm rate. This proves that samples over the USA are not enough to detect RFI robustly over Europe. Testing over Asia shows similar performance to Europe, where false alarm rates are very high. Samples from India also detect RFI with a very low false alarm rate as precision is 99.46% but missed some RFI cases which are apparent in 86.01% recall. This shows that the CNN-based DL algorithm finds it challenging to detect RFI correctly when it is trained and tested over different regions with possibly different RFI characteristics.

Finally, we train the DL-based RFI detection model over regions of Europe, Asia, and India separately and each learned model is tested over the USA region. Our first observation is that the detection model trained with Europe and Asia region performs very accurately, where all overall performance metrics are above 98.96%. Europe and Asia have most of the RFI cases globally, and possibly they include similar RFI cases to the USA region leading to high detection performance. We see that overall performance deteriorates slightly when it is trained

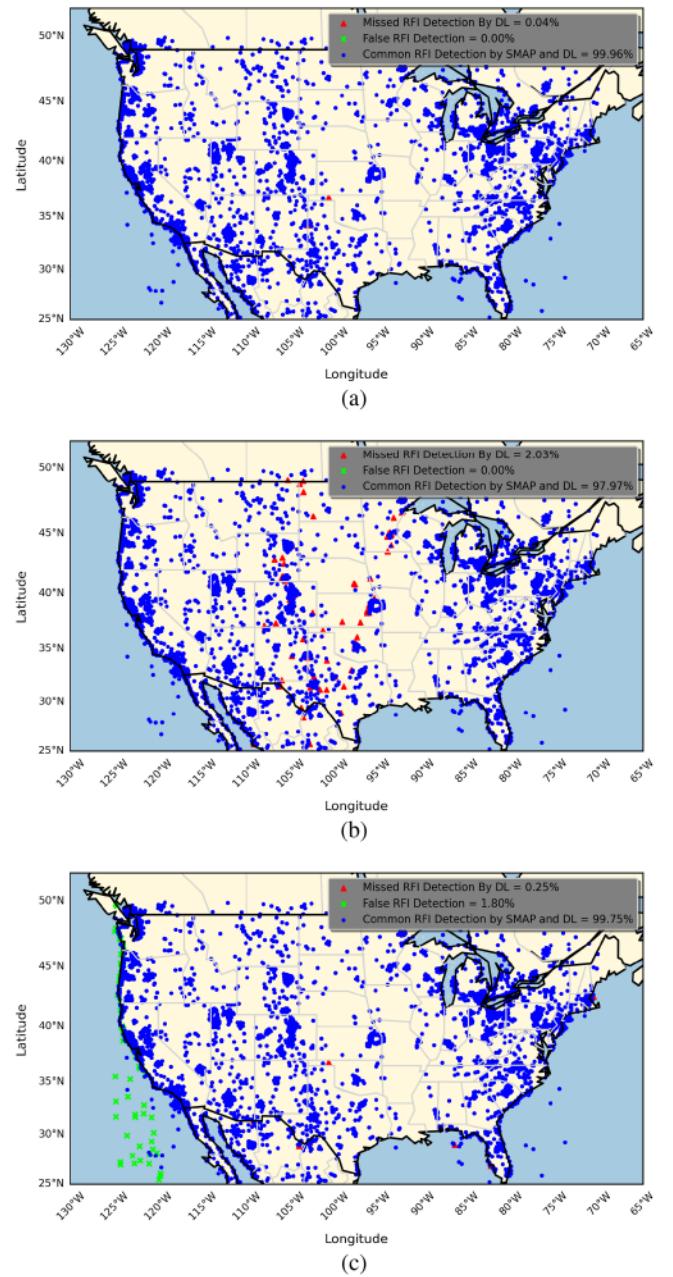


Fig. 10. RFI detection performance over USA trained with (a) Europe, (b) Asia, and (c) India.

with a smaller region such as India. The spatial distributions of all detection decisions are depicted in Fig. 10 to illustrate the performance of the DL-based model spatially. Fig. 10(a)–(c) demonstrates the experiment when the DL-based RFI detection model is trained with Europe, Asia, and India, respectively. In each figure, correct and missed detections as well as false alarms are shown spatially. While a high correct detection is demonstrated in Fig. 10(a) (trained with Europe), it is also observed that some missed RFI detections available in Fig. 10(b) (trained with Asia) and false alarms can be seen specifically on the western coastline in Fig. 10(c) (trained with India). This shows that high care should be given to designing the training dataset to make sure it has enough RFI variation to cover possible

RFI cases in the test region. When the model is trained with relatively smaller regions possibly with less RFI-type variation and a lower number of samples, the detection performance applied to another region might not be very high. Furthermore, special care must be exercised for the coastlines as the footprint is mixed with land and water. This, in turn, introduces systematic variable background on the spectrogram over time [18]. Hence, one can train the DL model just using samples from coastlines to create a detection framework specific to those regions that might have increased performance specifically in those regions.

V. SUMMARY AND CONCLUSION

In this study, it is demonstrated that radiometer RFI can be detected by using a single DL framework based on a CNN architecture. This study utilizes one coherent model to detect RFI-contaminated footprints using raw antenna counts, whereas the SMAP ground processing unit utilizes nine different algorithms to identify RFI. SMAP's Level 1 A data products are used to create the input spectrograms and level 1B quality flags are employed to dynamically label these spectrograms. DL model is trained and tested with vertical and horizontal polarized raw moments as well as third and fourth Stokes parameters. Four different CV techniques are evaluated to depict the robustness and flexibility of the model. As the dataset is heavily biased for one class ("No RFI" for this study), a single metric such as accuracy might not be enough performance indicator. Evaluations with performance metrics, such as precision, recall, and F1-score, are performed. Overall performance (F1-score) is above 99% for each dataset, which proves that the DL-based algorithm can be an attractive alternative in current and future earth observation satellites. Performance metrics of the time-based and regional analysis show that the DL-based RFI detection model is effective in diverse situations with a moderate number of samples for shorter time spans. Additional preprocessing steps on raw antenna count observations before they are utilized on DL training have the potential to mitigate receiver parameter drifts in time. In addition, this study is limited to detecting RFI with the entire spectrogram. Future studies can aim to develop DL-based RFI mitigation techniques based on this study's detection algorithm, which can also be extended into sub-band level RFI detection. In addition, studies on merging different polarization inputs under a maximum probability of detection framework may improve the proposed approach.

REFERENCES

- [1] P. Ferrazzoli and L. Guerriero, "Passive microwave remote sensing of forests: A model investigation," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 433–443, Mar. 1996.
- [2] M. M. Farhad, S. Biswas, M. A. S. Rafi, M. Kurum, and A. C. Gurbuz, "Design and implementation of a software defined radio-based radiometer operating from a small unmanned aircraft systems," in *Proc. IEEE USNC-URSI Radio Sci. Meeting*, 2022, pp. 17–18.
- [3] D. Entekhabi et al., "The soil moisture active passive (SMAP) mission," *Proc. IEEE*, vol. 98, no. 5, pp. 704–716, May 2010.
- [4] S. W. Ellingson, G. A. Hampson, and J. T. Johnson, "Characterization of L-band RFI and implications for mitigation techniques," *Int. Geosci. Remote Sens. Symp.*, vol. 3, pp. 1745–1747, 2003.
- [5] J. T. Johnson, G. A. Hampson, and S. W. Ellingson, "Design and demonstration of an interference suppressing microwave radiometer," *Int. Geosci. Remote Sens. Symp.*, vol. 3, pp. 1683–1686, 2004.
- [6] J. R. Piepmeier et al., "Radio-frequency interference mitigation for the soil moisture active passive microwave radiometer," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 761–775, Jan. 2014.
- [7] N. Majurec, J. Park, N. Niamsuwan, M. Frankford, and J. T. Johnson, "Airborne L-band RFI observations in the SMAPVEX08 campaign with the L-band interference suppressing radiometer," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009 pp. II-158–II-161.
- [8] N. Niamsuwan, J. T. Johnson, and S. W. Ellingson, "Examination of a simple pulse-blanking technique for radio frequency interference mitigation," *Radio Sci.*, vol. 40, pp. 1–11, 2005.
- [9] M. Steven Gross and S. Misra, "RFI detection and mitigation for microwave radiometry with an agile digital detector," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 3, pp. 694–706, Mar. 2006.
- [10] S. Misra, R. D. De Roo, and C. S. Ruf, "An improved radio frequency interference model: Reevaluation of the kurtosis detection algorithm performance under central-limit conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4565–4574, Nov. 2012.
- [11] B. Güner, N. Niamsuwan, and J. T. Johnson, "Performance study of a cross-frequency detection algorithm for pulsed sinusoidal RFI in microwave radiometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 7, pp. 2899–2908, Jul. 2010.
- [12] J. Piepmeier, P. Mohammed, G. De, A. Ed Kim, J. Peng, and C. Ruf, "Soil moisture active passive (SMAP) algorithm theoretical basis document (ATBD) SMAP calibrated, time-ordered brightness temperatures L1B TB data product," Jet Propulsion Lab., California Inst. Technol., Pasadena, CA, USA, 2014.
- [13] A. Bringer et al., "Properties of the RFI environment at 1400–1427 MHz as observed by the soil moisture active/passive mission microwave radiometer," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7259–7267, 2021.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [15] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, 521, pp. 436–444, 2015.
- [16] P. Artiemjew, A. Chojka, and J. Rapiński, "Deep learning for RFI artifact recognition in Sentinel-1 data," *Remote Sens.*, vol. 13, no. 1, pp. 1–7, 2020.
- [17] J. Akeret, C. Chang, A. Lucchi, and A. Refregier, "Radio frequency interference mitigation using deep convolutional neural networks," *Astron. Comput.*, vol. 18, pp. 35–39, 2017.
- [18] P. N. Mohammed and J. R. Piepmeier, "Microwave radiometer RFI detection using deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6398–6405, 2021.
- [19] "SMAP L1A radiometer time-ordered parsed telemetry, version 2," Nat. Snow Ice Data Center. [Online]. Available: <https://nsidc.org/data/SPL1AP/versions/2>
- [20] "SMAP L1B radiometer half-orbit time-ordered brightness temperatures, version 5," Nat. Snow Ice Data Center. [Online]. Available: <https://nsidc.org/data/SPL1BTB/versions/5>
- [21] *SMAP Handbook Soil Moisture Act. Passive Mapping Soil Moisture and freeze/thaw From Space*, Leuven, Belgium, 2014.
- [22] J. R. Piepmeier, D. G. Long, and E. G. Njoku, "Stokes antenna temperatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 2, pp. 516–527, Feb. 2008.
- [23] D. G. Long, M. J. Brodzik, and M. A. Hardman, "Enhanced-resolution SMAP brightness temperature image products," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4151–4163, Jul. 2019.
- [24] P. Mohammed, J. Peng, E. J. Kim, G. De Amici, J. Chaubell, and C. Ruf, "SMAP L1B radiometer half-orbit time-ordered brightness temperatures, version 5," Nat. Snow Ice Data Center, Boulder, CO, USA, 2020.
- [25] E. J. Kim, P. Mohammed, J. Peng, and C. Ruf, "SMAP L1A radiometer time-ordered parsed telemetry, version 2," Nat. Snow Ice Data Center, Boulder, CO, USA, 2015.
- [26] Y. Soldo, D. Le Vine, and P. de Mattheis, "Detection of residual "hot spots" in RFI-filtered SMAP data," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2935.
- [27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–14.
- [30] K. Weiss, M. Taghi Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, pp. 1–40, 2016.
- [31] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Appl. Soft Comput.*, vol. 97, 2020, Art. no. 105524.
- [32] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *Towards Data Sci.*, vol. 4, pp. 310–316, 2020.
- [33] A. F. Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*.
- [34] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Des. Implementation*, Savannah, GA, USA, vol. 2, pp. 265–283, 2016.
- [35] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2020.
- [36] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–15.
- [37] A. Senior, G. Heigold, A. Ranzato, and K. Yang, "An empirical study of learning rates in deep neural networks for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 6724–6728.
- [38] Y. Yao, L. Rosasco, and A. Caponnetto, "On early stopping in gradient descent learning," *Constructive Approximation*, vol. 26, pp. 289–315, 2007.
- [39] S. Yan et al., "An overview of overfitting and its solutions," *Proc. J. Phys., Conf. Ser.*, vol. 1168, 2019, Art. no. 022022.
- [40] D. Berrar, "Cross-validation," *Encyclopedia Database Syst.*, vol. 5, pp. 532–538, 2019.
- [41] B. Juba and H. S. Le, "Precision-recall versus accuracy and the role of large data sets," in *Proc. 33rd AAAI Conf. Artif. Intell., 31st Innov. Appl. Artif. Intell. Conf., 9th AAAI Symp. Educ. Adv. Artif. Intell.*, 2019, pp. 4039–4048.
- [42] D. M. W. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.
- [43] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, pp. 861–874, 2006.
- [44] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [45] S. Misra, P. N. Mohammed, B. Guner, C. S. Ruf, J. R. Piepmeyer, and J. T. Johnson, "Microwave radiometer radio-frequency interference detection algorithms: A comparative study," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3742–3754, Nov. 2009.
- [46] F. Pedregosa et al., "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.



Ahmed Manavi Alam (Graduate Student Member, IEEE) received the B.S. degree in electrical and electronic engineering from the Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in 2019. He is currently working toward the Ph.D. degree in electrical and computer engineering with Mississippi State University, Starkville, MS, USA.

He is currently working as a Research Assistant with the Information Processing and Sensing Laboratory. His research focus includes algorithm development of deep learning-based inverse problems and machine learning for remote sensing and physics-aware deep learning. He is a Student Member of the IEEE Geoscience and Remote Sensing Society.



Mehmet Kurum (Senior Member, IEEE) received the B.S. degree in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from George Washington University, Washington, DC, USA, in 2005 and 2009, respectively.

He held a Postdoctoral position with Hydrological Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA. He is currently an Associate Professor and Paul B. Jacob Endowed Chair in Electrical and Computer Engineering with Mississippi State University, where he is also the Co-Director of Information Processing and Sensing Laboratory. His current research focuses on recycling the radio spectrum to address the challenges of decreasing radio spectrum space for science while exploring entirely new microwave regions for land remote sensing.

Dr. Kurum is a Senior Member of the IEEE Geoscience and Remote Sensing Society and a Member of the U.S. National Committee for the International Union of Radio Science. He has been serving as an Associate Editor for *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING* and *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* since 2021. He was a recipient of the Leopold B. Felsen Award for excellence in electromagnetic in 2013, the International Union of Radio Science (URSI) Young Scientist Award in 2014, and the NSF CAREER award in 2022. He served as an Early Career Representative for the International URSI Commission F (Wave Propagation and Remote Sensing) from 2014 to 2021.



Ali C. Gurbuz (Senior Member, IEEE) received the B.S. degree in electrical engineering from Bilkent University, Ankara, Turkey, in 2003, and the M.S. and Ph.D. degrees from the Georgia Institute of Technology, Atlanta, GA, USA, in 2005 and 2008, both in electrical and computer engineering.

From 2003 to 2009, he researched compressive sensing based computational imaging problems with Georgia Tech. Between 2009 and 2017, he held faculty positions with TOBB University and the University of Alabama, where he pursued an active research program on the development of sparse signal representations, compressive sensing theory and applications, radar and sensor array signal processing, and machine learning. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, Mississippi State University, where he is the Co-Director of Information Processing and Sensing Laboratory.

Dr. Gurbuz was a recipient of The Best Paper Award for *Signal Processing Journal* in 2013 and the Turkish Academy of Sciences Best Young Scholar Award in electrical engineering in 2014, and NSF CAREER Award in 2021. He has served as an Associate editor for several journals such as *Digital Signal Processing*, *EURASIP Journal on Advances in Signal Processing*, and *Physical Communications*.